

1: (K) {20}
K
2: (K) {20}
K
3: (K) {20}
KK
4: (K) {20}
KKK
5: (K) {20}
K
6: (K) {20}
K
7: (K) {20}
K
8: (K) {20}
K
9: (K) {20}
K
10: (K) {20}
K
11: (K) {20}
K
12: (K) {20}
K
13: (K) {20}
K
14: (K) {20}
K
15: (K) {20}
K
16: (K) {20}
K
17: (K) {20}
K
18: (K) {20}
K
19: (K) {20}
K
20: (K) {20}
K
21: (K) {20}
K
22: (K) {20}
K
23: (K) {20}
K
24: (K) {20}
K

25: (K) {20}
K
26: (K) {20}
K
27: (K) {20}
K
28: (K) {20}
K
29: (K) {20}
K
30: (K) {20}
K
31: (K) {20}
K
32: (K) {20}
K
33: (K) {20}
K
34: (K) {20}
K
35: (K) {20}
K
36: (K) {20}
K
37: (K) {20}
K
38: (K) {20}
K
39: (K) {20}
K
40: (K) {20}
K
41: (K) {20}
K
42: (K) {20}
K
43: (K) {20}
K
44: (K) {20}
K
45: (K) {20}
K
46: (K) {20}
K
47: (K) {20}
K
48: (K) {20}
K
49: (K) {20}
K

13: KKKKK (K) {20} KKKKK
14: KKKKK (K) {20} KKKKK
15: KKKKK (K) {20} KKKKK
16: KKKKK (K) {20} KKKKK
17: KKKKK (K) {20} KKKKK
18: KKKKK (K) {20} KKKKK
19: KKKKK (K) {20} KKKKK
20: KKKKK (K) {20} KKKKK
21: KKKKK (K) {20} KKKKK
22: KKKKK (K) {20} KKKKK

AAW24865 ck: 2211 len: 40 ! Aaw24865 Bifunctional peptide I for binding

1

21: YEDES (R,K) {20,20} (K) {20} KKKKKKKKKKKKKKKKKKK

AAW24450 ck: 8137 len: 45 ! Aaw24450 Nucleic acid (NA) binding peptide

1

4: YKA (R,K) {20,20} (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
5: YKAK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
6: YKAKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
7: KAKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
8: AKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
9: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
10: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
11: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
12: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
13: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKK (K) {20} KKKKK
15: KKKKK (K) {20} KKKKK
16: KKKKK (K) {20} KKKKK
17: KKKKK (K) {20} KKKKK
18: KKKKK (K) {20} KKKKK
19: KKKKK (K) {20} KKKKK
20: KKKKK (K) {20} KKKKK
21: KKKKK (K) {20} KKKKK
22: KKKKK (K) {20} KKKKK
23: KKKKK (K) {20} KKKKK
24: KKKKK (K) {20} KKKKK

AAW21590 ck: 4875 len: 30 ! Aaw21590 Antibiotic potentiating peptide #2

1

1: (R,K) {20,20} (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
2: K (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
3: KK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
4: KKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
5: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
6: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
7: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
8: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
9: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
10: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK K
11: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK

AAW21591 ck: 5075 len: 434 ! Aaw21591 Antibiotic potentiating peptide #3

1

(R,K) {20,20}

8: KKKKK (K){20} KKKKK
9: KKKKK (K){20} KKKKK
10: KKKKK (K){20} KKKKK
11: KKKKK (K){20} KKKKK
12: KKKKK (K){20} KKKKK
13: KKKKK (K){20} KKKKK
14: KKKKK (K){20} KKKKK
15: KKKKK (K){20} KKKKK
16: KKKKK (K){20} KKKKK
17: KKKKK (K){20} KKKKK
18: KKKKK (K){20} KKKKK
19: KKKKK (K){20} KKKKK
20: KKKKK (K){20} KKKKK

AAW38816 ck: 8217 len: 42 ! Aaw38816 Delivery peptide used in peptide m
(R,K){20,20}
1: KKKKK (K){20} KKKKK
2: K KKKKK (K){20} KKKKK
3: KK KKKKK (K){20} KKKKK
4: KKK KKKKK (K){20} KKKKK
5: KKKK KKKKK (K){20} KKKKK
6: KKKKK KKKKK (K){20} KKKKK
7: KKKKK KKKKK (K){20} KKKKK
8: KKKKK KKKKK (K){20} KKKKK
9: KKKKK KKKKK (K){20} KKKKK
10: KKKKK KKKKK (K){20} KKKKK

11: KKKKK (K){20} KKKKK
12: KKKKK (K){20} KKKKK
13: KKKKK (K){20} KKKKK
14: KKKKK (K){20} KKKKK
15: KKKKK (K){20} KKKKK
16: KKKKK (K){20} KKKKK
17: KKKKK (K){20} KKKKK
18: KKKKK (K){20} KKKKK
19: KKKKK (K){20} KKKKK
20: KKKKK (K){20} KKKKK
21: KKKKK (K){20} KKKKK

AAW38817 ck: 1454 len: 43 ! Aaw38817 Delivery peptide used in peptide m
(R,K){20,20}
1: KKKKK (K){20} KKKKK
2: K KKKKK (K){20} KKKKK
3: KK KKKKK (K){20} KKKKK
4: KKK KKKKK (K){20} KKKKK
5: KKKKK KKKKK (K){20} KKKKK
6: KKKKK KKKKK (K){20} KKKKK
7: KKKKK KKKKK (K){20} KKKKK
8: KKKKK KKKKK (K){20} KKKKK
9: KKKKK KKKKK (K){20} KKKKK
10: KKKKK KKKKK (K){20} KKKKK
11: KKKKK KKKKK (K){20} KKKKK
12: KKKKK KKKKK (K){20} KKKKK

17: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK WK

AAW38813 ck: 8956 len: 39 ! Aaw38813 Delivery peptide used in peptide m
(R,K){20,20}
(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

2: K KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

5: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

AAW38814 ck: 1968 len: 40 ! Aaw38814 Delivery peptide used in peptide m
(R,K){20,20}
(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

2: K KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

5: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

AAW38815 ck: 5055 len: 41 ! Aaw38815 Delivery peptide used in peptide m
(R,K){20,20}
(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K){20,20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

5: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK (K){20}


```

14: KKKKK (K) {20}
15: KKKKK (K) {20}
16: KKKKK (K) {20}

AAW38893 ck: 9249 len: 39 ! Aaw38893 Delivery peptide used in peptide m
(R,K) {20,20}
1: KKKKK
2: K KKKKK
3: KK KKKKK
4: KKK KKKKK
5: KKKK KKKKK
6: KKKKK KKKKK
7: KKKKK KKKKK
8: KKKKK KKKKK
9: KKKKK KKKKK
10: KKKKK KKKKK
11: KKKKK KKKKK
12: KKKKK KKKKK
13: KKKKK KKKKK
14: KKKKK KKKKK
15: KKKKK KKKKK
16: KKKKK KKKKK
17: KKKKK KKKKK

AAW38894 ck: 2269 len: 40 ! Aaw38894 Delivery peptide used in peptide m
(R,K) {20,20}
1: KKKKK
(K) {20}

2: K KKKKK
3: KK KKKKK
4: KKK KKKKK
5: KKKK KKKKK
6: KKKKK KKKKK
7: KKKKK KKKKK
8: KKKKK KKKKK
9: KKKKK KKKKK
10: KKKKK KKKKK
11: KKKKK KKKKK
12: KKKKK KKKKK
13: KKKKK KKKKK
14: KKKKK KKKKK
15: KKKKK KKKKK
16: KKKKK KKKKK
17: KKKKK KKKKK

AAW38895 ck: 5364 len: 41 ! Aaw38895 Delivery peptide used in peptide m
(R,K) {20,20}
1: KKKKK
2: K KKKKK
3: KK KKKKK
4: KKK KKKKK
5: KKKK KKKKK
6: KKKKK KKKKK

```


21: KKKK KKKKKKKKKKKKKKKKKKK KK

AAW3888 ck: 5274 len: 34 ! Aaw3888 Delivery peptide used in peptide m

1

- 1: (R,K){20,20}
(K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 2: K (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 3: KK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 4: KKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 5: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 6: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 7: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 8: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 9: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 11: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 12: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK

AAW3854 ck: 1496 len: 43 ! Aaw3854 Delivery peptide used in peptide m

1

- 1: (R,K){20,20}
(K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 2: K (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 3: KK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 4: KKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 5: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 6: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 7: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 8: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 9: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK

- 10: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 11: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 12: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 13: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 14: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 15: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 16: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 17: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 18: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 19: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 20: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 21: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 22: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK

AAW3889 ck: 7919 len: 35 ! Aaw3889 Delivery peptide used in peptide m

1

- 1: (R,K){20,20}
(K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 2: K (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 3: KK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 4: KKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 5: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 6: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 7: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 8: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 9: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK
- 10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKK

AAW38885 ck: 7789 len: 31 ! Aaw38885 Delivery peptide used in peptide m

1: (R,K){20,20}
(K){20}

2: K (K){20}

3: KK (K){20}

4: KKK (K){20}

5: KKKK (K){20}

6: KKKKK (K){20}

7: KKKKKK (K){20}

8: KKKKKKK (K){20}

9: KKKKKKKK (K){20}

AAW38851 ck: 2007 len: 40 ! Aaw38851 Delivery peptide used in peptide m

1: (R,K){20,20}
(K){20}

2: K (K){20}

3: KK (K){20}

4: KKK (K){20}

5: KKKK (K){20}

6: KKKKK (K){20}

7: KKKKKK (K){20}

8: KKKKKKK (K){20}

9: KKKKKKKK (K){20}

10: KKKKKKKKK (K){20}

11: KKKKKKKKKK (K){20}

12: KKKKKKKKKKK (K){20}

13: KKKKKKKKKKKK (K){20}

14: KKKKK (K){20}

15: KKKKKK (K){20}

16: KKKKKKK (K){20}

17: KKKKKKKK (K){20}

18: KKKKKKKKK (K){20}

19: KKKKKKKKKK (K){20}

AAW38886 ck: 209 len: 32 ! Aaw38886 Delivery peptide used in peptide m

1: (R,K){20,20}
(K){20}

2: K (K){20}

3: KK (K){20}

4: KKK (K){20}

5: KKKK (K){20}

6: KKKKK (K){20}

7: KKKKKK (K){20}

8: KKKKKKK (K){20}

9: KKKKKKKK (K){20}

10: KKKKKKKKK (K){20}

AAW38852 ck: 5095 len: 41 ! Aaw38852 Delivery peptide used in peptide m

1: (R,K){20,20}
(K){20}

2: K (K){20}

3: KK (K){20}

4: KKK (K){20}

5: KKKK (K){20}

6: KKKK KKKKKKKKKKKKKKKKKKK KVTK
(K) {20}
7: KKKK KKKKKKKKKKKKKKKKKKK VTK

AAW38849 ck: 6056 len: 38 ! Aaw38849 Delivery peptide used in peptide m

1

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
(K) {20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAW38884 ck: 5444 len: 30 ! Aaw38884 Delivery peptide used in peptide m

1

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAW38850 ck: 8994 len: 39 ! Aaw38850 Delivery peptide used in peptide m

1

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

5: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

5: KKKK KKKK{20} VTK

AAW38847 ck: 405 len: 36 ! Aaw38847 Delivery peptide used in peptide m

1

- 1: KKKK{20,20}
- 2: K KKKK{20}
- 3: KK KKKK{20}
- 4: KKK KKKK{20}
- 5: KKKK KKKK{20}
- 6: KKKK KKKK{20}
- 7: KKKK KKKK{20}
- 8: KKKK KKKK{20}
- 9: KKKK KKKK{20}
- 10: KKKK KKKK{20}
- 11: KKKK KKKK{20}
- 12: KKKK KKKK{20}
- 13: KKKK KKKK{20}
- 14: KKKK KKKK{20}
- 15: KKKK KKKK{20} KK

AAW38882 ck: 979 len: 28 ! Aaw38882 Delivery peptide used in peptide m

1

- 1: KKKK{20,20}
- 2: K KKKK{20}
- 3: KK KKKK{20}
- 4: KKK KKKK{20}
- 5: KKKK KKKK{20}
- 6: KKKK KKKK{20} VTK

AAW38848 ck: 3193 len: 37 ! Aaw38848 Delivery peptide used in peptide m

1

- 1: KKKK{20,20}
- 2: K KKKK{20}
- 3: KK KKKK{20}
- 4: KKK KKKK{20}
- 5: KKKK KKKK{20}
- 6: KKKK KKKK{20}
- 7: KKKK KKKK{20}
- 8: KKKK KKKK{20}
- 9: KKKK KKKK{20}
- 10: KKKK KKKK{20}
- 11: KKKK KKKK{20}
- 12: KKKK KKKK{20}
- 13: KKKK KKKK{20}
- 14: KKKK KKKK{20}
- 15: KKKK KKKK{20} KK
- 16: KKKK KKKK{20} KK

AAW38883 ck: 3174 len: 29 ! Aaw38883 Delivery peptide used in peptide m

1

- 1: KKKK{20,20}
- 2: K KKKK{20}
- 3: KK KKKK{20}
- 4: KKK KKKK{20}
- 5: KKKK KKKK{20} KVT

8: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKK
9: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKK
10: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKK
11: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKK

AAW38806 ck: 9972 len: 32 ! Aaw38806 Delivery peptide used in peptide m
(R,K){20,20}
(K){20}

1: KKKKK (R,K){20,20} KKKKK
2: K KKKKK (K){20} KKKKK
3: KK KKKKK (K){20} KKKKK
4: KKK KKKKK (K){20} KKKKK
5: KKK KKKKK (K){20} KKKKK
6: KKKKK KKKKK (K){20} KKKKK
7: KKKKK KKKKK (K){20} KKKKK
8: KKKKK KKKKK (K){20} KKKKK
9: KKKKK KKKKK (K){20} KKKKK
10: KKKKK KKKKK (K){20} KKKKK
11: KKKKK KKKKK (K){20} KKKKK

AAW38807 ck: 2459 len: 33 ! Aaw38807 Delivery peptide used in peptide m
(R,K){20,20}
(K){20}

1: KKKKK (R,K){20,20} KKKKK
2: K KKKKK (K){20} KKKKK
3: KK KKKKK (K){20} KKKKK
4: KKK KKKKK (K){20} KKKKK
5: KKK KKKKK (K){20} KKKKK
6: KKKKK KKKKK (K){20} KKKKK
7: KKKKK KKKKK (K){20} KKKKK

8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
9: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKK
10: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKK
11: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKK
12: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK WK

AAW38808 ck: 5021 len: 34 ! Aaw38808 Delivery peptide used in peptide m
(R,K){20,20}
(K){20}

1: KKKKK (R,K){20,20} KKKKK
2: K KKKKK (K){20} KKKKK
3: KK KKKKK (K){20} KKKKK
4: KKK KKKKK (K){20} KKKKK
5: KKK KKKKK (K){20} KKKKK
6: KKKKK KKKKK (K){20} KKKKK
7: KKKKK KKKKK (K){20} KKKKK
8: KKKKK KKKKK (K){20} KKKKK
9: KKKKK KKKKK (K){20} KKKKK
10: KKKKK KKKKK (K){20} KKKKK
11: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKK
12: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKK
13: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK WK

AAW38881 ck: 8859 len: 27 ! Aaw38881 Delivery peptide used in peptide m
(R,K){20,20}
(K){20}

1: KKKKK (R,K){20,20} KKKKK
2: K KKKKK (K){20} KKKKK
3: KK KKKKK (K){20} KKKKK
4: KKK KKKKK (K){20} KKKKK

```

1
1: (R,K){20,20}
(K){20}
2: K (K){20}
3: KK (K){20}
4: KKK (K){20}
5: KKKK (K){20}
AAW38801 ck: 8662 len: 27 ! Aaw38801 Delivery peptide used in peptide m
1: (R,K){20,20}
(K){20}
2: K (K){20}
3: KK (K){20}
4: KKK (K){20}
5: KKKK (K){20}
6: KKKKK (K){20}
AAW38802 ck: 774 len: 28 ! Aaw38802 Delivery peptide used in peptide m
1: (R,K){20,20}
(K){20}
2: K (K){20}
3: KK (K){20}
4: KKK (K){20}
5: KKKK (K){20}
6: KKKKK (K){20}
AAW38803 ck: 2961 len: 29 ! Aaw38803 Delivery peptide used in peptide m
1: (R,K){20,20}
(K){20}
2: K (K){20}

```

```

3: KK (K){20}
4: KKK (K){20}
5: KKKK (K){20}
6: KKKKK (K){20}
7: KKKKKK (K){20}
8: KKKKKKK (K){20}
AAW38804 ck: 5223 len: 30 ! Aaw38804 Delivery peptide used in peptide n
1: (R,K){20,20}
(K){20}
2: K (K){20}
3: KK (K){20}
4: KKK (K){20}
5: KKKK (K){20}
6: KKKKK (K){20}
7: KKKKKK (K){20}
8: KKKKKKK (K){20}
9: KKKKKKKK (K){20}
AAW38805 ck: 7560 len: 31 ! Aaw38805 Delivery peptide used in peptide n
1: (R,K){20,20}
(K){20}
2: K (K){20}
3: KK (K){20}
4: KKK (K){20}
5: KKKK (K){20}
6: KKKKK (K){20}
7: KKKKKK (K){20}
8: KKKKKKK (K){20}
9: KKKKKKKK (K){20}

```


1
 1: (R,K){20,20}
 (K){20}
 2: K
 (K){20}
 3: K
 (K){20}
 4: K
 (K){20}
 5: K
 (K){20}
 6: K
 (K){20}
 7: K
 (K){20}
 8: K
 (K){20}

AAW38841 ck: 5252 len: 30 ! Aaw38841 Delivery peptide used in peptide

1
 1: (R,K){20,20}
 (K){20}
 2: K
 (K){20}
 3: K
 (K){20}
 4: K
 (K){20}
 5: K
 (K){20}
 6: K
 (K){20}
 7: K
 (K){20}
 8: K
 (K){20}
 9: K
 (K){20}

AAW38842 ck: 7590 len: 31 ! Aaw38842 Delivery peptide used in peptide

1
 1: (R,K){20,20}
 (K){20}
 2: K
 (K){20}
 3: K
 (K){20}
 4: K
 (K){20}

5: K
 (K){20}
 6: K
 (K){20}
 7: K
 (K){20}
 8: K
 (K){20}
 9: K
 (K){20}
 10: K
 (K){20}

AAW38877 ck: 1129 len: 23 ! Aaw38877 Delivery peptide used in peptide

1
 1: (R,K){20,20}
 (K){20}

AAW38843 ck: 3 len: 32 ! Aaw38843 Delivery peptide used in peptide

1
 1: (R,K){20,20}
 (K){20}
 2: K
 (K){20}
 3: K
 (K){20}
 4: K
 (K){20}
 5: K
 (K){20}
 6: K
 (K){20}
 7: K
 (K){20}
 8: K
 (K){20}
 9: K
 (K){20}
 10: K
 (K){20}
 11: K
 (K){20}

AAW38878 ck: 2949 len: 24 ! Aaw38878 Delivery peptide used in peptide

1
 1: (R,K){20,20}
 (K){20}
 2: K
 (K){20}

AAW38844 ck: 2491 len: 33 ! Aaw38844 Delivery peptide used in peptide

! FINDPATTERNS on geneseqp:* allowing 0 mismatches

! 1 (R,K){20,20} January 30, 2004 07:18 ..

AAP20159 ck: 5750 len: 20 ! Aap20159 Sequence of lysine polymer. 8/2002

1: (R,K){20,20}
(K){20}
XXXXXXXXXXXXXXXXXXXX

AAP61030 ck: 9157 len: 898 ! Aap61030 Entire coded sequence from plasmid

873: KNITW XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(K){20}
874: NITWK XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(K){20}
875: ITWKK XXXXXXXXXXXXXXXXXXXXXXXX KKKK
(K){20}
876: TWKKK XXXXXXXXXXXXXXXXXXXXXXXX KKK
(K){20}
877: WKKKK XXXXXXXXXXXXXXXXXXXXXXXX KK
(K){20}
878: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX K
(K){20}
879: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX

AAP61056 ck: 2017 len: 899 ! Aap61056 Translation of plasmid PAU157 encd

873: KNITW XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(K){20}
874: NITWK XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(K){20}
875: ITWKK XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(K){20}
876: TWKKK XXXXXXXXXXXXXXXXXXXXXXXX KKKK
(K){20}
877: WKKKK XXXXXXXXXXXXXXXXXXXXXXXX KKK
(K){20}
878: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX KK
(K){20}
879: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX K
(K){20}
880: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX

AAP61082 ck: 7915 len: 898 ! Aap61082 Complete translation of plasmid PA

873: KNITW XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(R,K){20,20}
(K){20}
874: NITWK XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(K){20}

875: ITWKK XXXXXXXXXXXXXXXXXXXXXXXX KKKK
(K){20}
876: TWKKK XXXXXXXXXXXXXXXXXXXXXXXX KKK
(K){20}
877: WKKKK XXXXXXXXXXXXXXXXXXXXXXXX KK
(K){20}
878: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX K
(K){20}
879: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX

AAR29580 ck: 4341 len: 657 ! Aar29580 FMR-1 gene product. 3/2003

19: RRRRP XXXXXXXXXXXXXXXXXXXXXXXX RRRRL
(R,K){20,20}
(R){20}
20: RRRRP XXXXXXXXXXXXXXXXXXXXXXXX RRLG
(R){20}
21: RRRRP XXXXXXXXXXXXXXXXXXXXXXXX RRLGL
(R){20}
22: RRRRP XXXXXXXXXXXXXXXXXXXXXXXX RLGLE
(R){20}
23: RRRRP XXXXXXXXXXXXXXXXXXXXXXXX LGLER
(R){20}

AAW03642 ck: 9623 len: 116 ! Aaw03642 Human cannabinoid GPR N-terminal s

34: QYEDI XXXXXXXXXXXXXXXXXXXXXXXX KSPFQ
(R,K){20,20}
(K){20}
35: YEDIK XXXXXXXXXXXXXXXXXXXXXXXX SPFOE
(K){20}

AAW38839 ck: 801 len: 28 ! Aaw38839 Delivery peptide used in peptide m

1: XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(R,K){20,20}
(K){20}
2: K XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(K){20}
3: KK XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(K){20}
4: KKK XXXXXXXXXXXXXXXXXXXXXXXX KKKKK
(K){20}
5: KKKK XXXXXXXXXXXXXXXXXXXXXXXX KKKK
(K){20}
6: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX KKK
(K){20}
7: KKKKK XXXXXXXXXXXXXXXXXXXXXXXX KK

AAW38840 ck: 2989 len: 29 ! Aaw38840 Delivery peptide used in peptide m

224: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKR
 (K) {20}
 225: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKRH
 (K) {20}
 226: KKKKK KKKKKKKKKKKKKKKKKKKKK KKRHH
 (K) {20}
 227: KKKKK KKKKKKKKKKKKKKKKKKKKK KKRHH

Q64075 ck: 8048 len: 215 ! Q64075 rattus sp. nucleoporin p62 homolog

(R,K) {20,20}
 (K) {20}

35: CEFLE KKKKKKKKKKKKKKKKKKKKK KKTGD
 (K) {20}
 36: EFLEK KKKKKKKKKKKKKKKKKKKKK KTGDN
 (K) {20}
 37: FLEKK KKKKKKKKKKKKKKKKKKKKK TGDNA

Q9D5G1 ck: 9388 len: 169 ! Q9D5G1 mus musculus (mouse). adult male tes

(R,K) {20,20}
 (R) {20}

117: VQLRG RRRRRRRRRRRRRRRRRRRR RKKKK
 (R) {20}
 118: QLGR RRRRRRRRRRRRRRRRRRRR RKKKE
 (R) {20}
 119: LRGR RRRRRRRRRRRRRRRRRRRR KKEEE
 (R,K) {20}
 120: RGR RRRRRRRRRRRRRRRRRRRR KKEEE
 (R,K) {20}
 121: GRRR RRRRRRRRRRRRRRRRRRRR KEEEE
 (R,K) {20}
 122: RRRR RRRRRRRRRRRRRRRRRRRR EEEEE

Q35807 ck: 7510 len: 129 ! Q35807 rattus norvegicus (rat). microvascul

(R,K) {20,20}
 (K) {20}

85: VLLAS KKKKKKKKKKKKKKKKKKKKK KIKWE
 (K) {20}
 86: LLASK KKKKKKKKKKKKKKKKKKKKK IKWEG

Q8BXG9 ck: 5434 len: 115 ! Q8BXG9 mus musculus (mouse). hypothetical a

(R,K) {20,20}
 (R,K) {20}

40: IIIII RRRRRRRRRRRRRRRRRRRR KETGS
 (R,K) {20}
 41: IIIIR RRRRRRRRRRRRRRRRRRRR ETGSH

Q8BHV2 ck: 8958 len: 154 ! Q8BHV2 mus musculus (mouse). weakly similar

(R,K) {20,20}
 (R) {20}

42: RRGEE RRRRRRRRRRRRRRRRRRRR RRRRR

(R) {20}
 43: RGEER RRRRRRRRRRRRRRRRRRRR RRRRR
 (R) {20}
 44: GEERR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R) {20}
 45: EERRR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R) {20}
 46: ERRRR RRRRRRRRRRRRRRRRRRRR RRRRR
 (R) {20}
 47: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRK
 (R) {20}
 48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRKR
 (R) {20}
 49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRE
 (R) {20}
 50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRER
 (R,K) {20}
 51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRER
 (R,K) {20}
 52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRER
 (R,K) {20}
 53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRER

Databases searched:

SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
 SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

Total finds: 348
 Total length: 305,079,309
 Total sequences: 958,388
 CPU time: 11:25.82

176: KKKKK (K) {20} KKKKK
177: KKKKK (K) {20} KKKKK
178: KKKKK (K) {20} KKKKK
179: KKKKK (K) {20} KKKKK
180: KKKKK (K) {20} KKKKK
181: KKKKK (K) {20} KKKKK
182: KKKKK (K) {20} KKKKK
183: KKKKK (K) {20} KKKKK
184: KKKKK (K) {20} KKKKK
185: KKKKK (K) {20} KKKKK
186: KKKKK (K) {20} KKKKK
187: KKKKK (K) {20} KKKKK
188: KKKKK (K) {20} KKKKK
189: KKKKK (K) {20} KKKKK
190: KKKKK (K) {20} KKKKK
191: KKKKK (K) {20} KKKKK
192: KKKKK (K) {20} KKKKK
193: KKKKK (K) {20} KKKKK
194: KKKKK (K) {20} KKKKK
195: KKKKK (K) {20} KKKKK
196: KKKKK (K) {20} KKKKK
197: KKKKK (K) {20} KKKKK
198: KKKKK (K) {20} KKKKK
199: KKKKK (K) {20} KKKKK

200: KKKKK (K) {20} KKKKK
201: KKKKK (K) {20} KKKKK
202: KKKKK (K) {20} KKKKK
203: KKKKK (K) {20} KKKKK
204: KKKKK (K) {20} KKKKK
205: KKKKK (K) {20} KKKKK
206: KKKKK (K) {20} KKKKK
207: KKKKK (K) {20} KKKKK
208: KKKKK (K) {20} KKKKK
209: KKKKK (K) {20} KKKKK
210: KKKKK (K) {20} KKKKK
211: KKKKK (K) {20} KKKKK
212: KKKKK (K) {20} KKKKK
213: KKKKK (K) {20} KKKKK
214: KKKKK (K) {20} KKKKK
215: KKKKK (K) {20} KKKKK
216: KKKKK (K) {20} KKKKK
217: KKKKK (K) {20} KKKKK
218: KKKKK (K) {20} KKKKK
219: KKKKK (K) {20} KKKKK
220: KKKKK (K) {20} KKKKK
221: KKKKK (K) {20} KKKKK
222: KKKKK (K) {20} KKKKK
223: KKKKK (K) {20} KKKKK

127: (K) {20} XXXXX
128: (K) {20} XXXXX
129: (K) {20} XXXXX
130: (K) {20} XXXXX
131: (K) {20} XXXXX
132: (K) {20} XXXXX
133: (K) {20} XXXXX
134: (K) {20} XXXXX
135: (K) {20} XXXXX
136: (K) {20} XXXXX
137: (K) {20} XXXXX
138: (K) {20} XXXXX
139: (K) {20} XXXXX
140: (K) {20} XXXXX
141: (K) {20} XXXXX
142: (K) {20} XXXXX
143: (K) {20} XXXXX
144: (K) {20} XXXXX
145: (K) {20} XXXXX
146: (K) {20} XXXXX
147: (K) {20} XXXXX
148: (K) {20} XXXXX
149: (K) {20} XXXXX
150: (K) {20} XXXXX

151: (K) {20} XXXXX
152: (K) {20} XXXXX
153: (K) {20} XXXXX
154: (K) {20} XXXXX
155: (K) {20} XXXXX
156: (K) {20} XXXXX
157: (K) {20} XXXXX
158: (K) {20} XXXXX
159: (K) {20} XXXXX
160: (K) {20} XXXXX
161: (K) {20} XXXXX
162: (K) {20} XXXXX
163: (K) {20} XXXXX
164: (K) {20} XXXXX
165: (K) {20} XXXXX
166: (K) {20} XXXXX
167: (K) {20} XXXXX
168: (K) {20} XXXXX
169: (K) {20} XXXXX
170: (K) {20} XXXXX
171: (K) {20} XXXXX
172: (K) {20} XXXXX
173: (K) {20} XXXXX
174: (K) {20} XXXXX
175: (K) {20} XXXXX


```

1
446: RVGKK (K){20}
447: VQKKK (K){20}
Q8S7D3 ck: 6479 len: 80 ! Q8s7d3 oryza sativa (rice). hypothetical 9
1
48: VIHLN (R,K){20,20}
49: IHLN (K){20}
50: HLN (K){20}
51: LK (K){20}
52: DK (K){20}
53: KKKK (K){20}
54: KKKK (K){20}
55: KKKK (K){20}
56: KKKK (K){20}
57: KKKK (K){20}
Q8LQ6 ck: 3239 len: 113 ! Q8lq6 oryza sativa (japonica cultivar-group)
1
10: SLEHT (R,K){20,20}
11: LEHT (K){20}
12: EHT (K){20}
13: HT (K){20}
14: KKKK (K){20}
15: KKKK (K){20}
16: KKKK (K){20}
17: KKKK (K){20}
18: KKKK (R,K){20}
Q9LGZ9 ck: 6094 len: 260 ! Q9lgz9 arabidopsis thaliana (mouse-ear cress)

```

```

1
6: MDRCI (R,K){20,20}
7: DRCIR (K){20}
8: RCIRK (K){20}
9: CIRCK (K){20}
10: IRCKK (K){20}
11: RCKCK (K){20}
12: KCKCK (K){20}
13: KCKCK (K){20}
14: KCKCK (K){20}
15: KCKCK (K){20}
16: KCKCK (K){20}
17: KCKCK (K){20}
18: KCKCK (K){20}
19: KCKCK (K){20}
20: KCKCK (K){20}
21: KCKCK (K){20}
22: KCKCK (K){20}
23: KCKCK (K){20}
24: KCKCK (K){20}
25: KCKCK (K){20}
26: KCKCK (K){20}
27: KCKCK (K){20}
28: KCKCK (K){20}
29: KCKCK (K){20}

```

22: KRKKK (K) {20} RRRRR
23: KRKKK (R,K) {20} RRRRR
24: KRKKK (R,K) {20} RRRRG
25: RKKKK (R,K) {20} RRRRR
26: KKKKK (R,K) {20} RRRRR
27: KKKKK (R,K) {20} RRRRR
28: KKKKK (R,K) {20} RRRRR
Q9HC48 ck: 7602 len: 667 ! Q9hc48 homo sapiens (human). ctcl tumor and
(R,K) {20,20}
(R,K) {20}
648: GDKTD RKKKKK RKKKKK RKKKKK
Q9H5V6 ck: 379 len: 168 ! Q9h5v6 homo sapiens (human). hypothetical P
(R,K) {20,20}
(K) {20}
140: VREWE KKKKKK KKKKKK KKKKK
141: REMEK KKKKKK KKKKKK KKKKK
142: EWEKK KKKKKK KKKKKK KKKKK
143: WEKKK KKKKKK KKKKKK KKKKK
144: EKKKK KKKKKK KKKKKK KKKKK
145: KKKKK KKKKKK KKKKKK KKKKK
146: KKKKK KKKKKK KKKKKK KKK
147: KKKKK KKKKKK KKKKKK KK
148: KKKKK KKKKKK KKKKKK K
149: KKKKK KKKKKK KKKKKK KKK

Q8SWR7 ck: 2372 len: 515 ! Q8swr7 drosophila melanogaster (fruit fly).
(R,K) {20,20}
(K) {20}
493: PTDYI KKKKKK KKKKKK KKKKK
494: TDYIK KKKKKK KKKKKK KKK

495: DYIKK (K) {20} KKKKKK KKKKKK K
496: YIKKK (K) {20} KKKKKK KKKKKK KKKKKK
Q8T2U7 ck: 8768 len: 791 ! Q8t2u7 dictyostelium discoideum (slime mol
(R,K) {20,20}
(K) {20}
769: EIEKE KKKKKK KKKKKK KKKKK KKK
770: IEIEK KKKKKK KKKKKK KKKKKK EI
(K) {20}
Q8I247 ck: 5951 len: 206 ! Q8i247 plasmodium falciparum (isolate 3d7)
(R,K) {20,20}
(K) {20}
185: KKIIFL KKKKKK KKKKKK KKKKK KT
(K) {20}
186: KIIFLK KKKKKK KKKKKK KKKKK T
Q95LV6 ck: 7515 len: 531 ! Q95lv6 macaca fascicularis (crab eating ma
(R,K) {20,20}
(K) {20}
502: YKNS KKKKKK KKKKKK KKKKKK KKKKK
(K) {20}
503: KGNSK KKKKKK KKKKKK KKKKK KKKKK
(K) {20}
504: GNSKK KKKKKK KKKKKK KKKKK KKKKK
(K) {20}
505: NSKKK KKKKKK KKKKKK KKKKK KKKKK
(K) {20}
506: SKKKK KKKKKK KKKKKK KKKKK KKKKK
(K) {20}
507: KKKKK KKKKKK KKKKKK KKKKK KKKKK
(K) {20}
508: KKKKK KKKKKK KKKKKK KKKKK KKKKK
(K) {20}
509: KKKKK KKKKKK KKKKKK KKKKK KKK
(K) {20}
510: KKKKK KKKKKK KKKKKK KKKKK KK
(K) {20}
511: KKKKK KKKKKK KKKKKK KKKKK K
(K) {20}
512: KKKKK KKKKKK KKKKKK KKKKK KKK

Q9LXR2 ck: 4143 len: 517 ! Q9lxr2 arabidopsis thaliana (mouse-ear cre
(R,K) {20,20}
(K) {20}
444: FERVG KKKKKK KKKKKK KKKKK KKKIR
(K) {20}
445: ERVGK KKKKKK KKKKKK KKKKK KKKIR

! FINDPATTERNS on swp:* allowing 0 mismatches

! 1 (R,K){20,20}

January 30, 2004 07:00 ..

Q12444 ck: 1384 len: 126 ! Q12444 saccharomyces cerevisiae (baker's yeast)

(R,K){20,20}
(R,K){20}

53: RKERT RKRRRRRRRRRRRRRRRRR KRSPR

(R,K){20}

54: KRTR RKRRRRRRRRRRRRRRR RSPRK

(R,K){20}

55: RTTR RKRRRRRRRRRRRRRRR SPRKR

Q9P529 ck: 291 len: 128 ! Q9P529 neurospora crassa. hypothetical 15.2

(R,K){20,20}
(K){20}

71: KRNQ KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

72: RNQK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

73: RNQK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

74: NQKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

75: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

76: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

77: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

78: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

79: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

80: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

81: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

82: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

83: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

Q9NT34 ck: 7330 len: 380 ! Q9NT34 homo sapiens (human). hypothetical P

(R,K){20,20}
(K){20}

355: NLLQ KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

356: LLLQ KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

357: LLQK KKKKKKKKKKKKKKKKKKK KKKKK

358: LQKK (K){20} KKKKKKKKKKKKKKKKKKK KKK

(K){20}

359: QKKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

360: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

361: KKKKK KKKKKKKKKKKKKKKKKKK

Q9H6Q7 ck: 3351 len: 720 ! Q9H6Q7 homo sapiens (human). hypothetical 1

(R,K){20,20}

(K){20}

692: IVSIS KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

693: VSISK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

694: SISKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

695: ISKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

696: SKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

697: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

698: KKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

699: KKKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

700: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

701: KKKKK KKKKKKKKKKKKKKKKKKK

Q8N6F0 ck: 9898 len: 55 ! Q8N6F0 homo sapiens (human). similar to loc

(R,K){20,20}

(R,K){20}

14: RRGGRG KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

15: RRGK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

16: GRGK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

17: RGGK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

18: GKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

A:Map position: 15R
 C:Keywords: transmembrane protein
 F:3-19/Domain: transmembrane #status predicted <TM1>
 F:107-123/Domain: transmembrane #status predicted <TM2>
 S58321 Length: 126 January 30, 2004 07:58 Type: P Check: 1384 ..
 1 MQMLIPQRL LILNPILMMK RKRKRKRKR RERETMMKIP RILKKLRKR
 51 RTRRKRKRKR KRRRKRKRKR RKRSPRKR KRRNKDAFYI LIISDFSRSL
 101 LFGFRKFSII IQCLTYVFSFH ILFHNL

```

!!SEQUENCE LIST 1.0
! FINDPATTERNS on swp:* allowing 0 mismatches
!      1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {
SW:HRA1_MOUSE      ck: 7206 len: 480 finds: 1 | Q9rl18 mus musculus (mouse). s
SP_BA:Q9KHD9      ck: 1697 len: 262 finds: 1 | Q9khd9 streptomyces griseus su
SP_PL:Q8RUS6      ck: 3612 len: 107 finds: 1 | Q8rus6 oryza sativa (japonica
SP_RO:Q9QZK6      ck: 7594 len: 480 finds: 1 | Q9qzk6 mus musculus (mouse). i
SP_RO:Q91WS3      ck: 7882 len: 480 finds: 1 | Q91ws3 mus musculus (mouse). f
SP_RO:Q9QZK5      ck: 8689 len: 480 finds: 1 | Q9qzk5 rattus norvegicus (rat)
\\End of list

```

```

Databases searched:
SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

Total finds: 6
Total length: 305,079,309
Total sequences: 958,388
CPU time: 16:49.45

```

```

!!AA SEQUENCE 1.0
ID HRA1_MOUSE STANDARD; PRT; 480 AA.
AC Q9RII8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease HTRA1 precursor (EC 3.4.21.-).
GN PRS11 OR HTRA1 OR HTRA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RA Oka C., Soma A., Kanda H., Kawauchi M.;
RT "The role of murine serine protease HTRA in osteogenesis."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Protease that regulate the availability of IGFs by
CC cleaving IGF-binding proteins (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- SIMILARITY: Contains 1 IGFBP domain.
CC -!- SIMILARITY: Contains 1 Kazal-like domain.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF12994; AAD49422.1; -.
DR MEROPS; S01.277; -.
DR MGD; MGI:1929076; Prs11.
DR InterPro; IPR000867; Insl_gro_fac_pr.
DR InterPro; IPR002350; kazal.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00050; kazal; 1.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00121; IB; 1.
DR SMART; SM00280; KAZAL; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Growth factor binding; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 480 SERINE PROTEASE HTRA1.
FT DOMAIN 37 94 IGFBP.
FT DOMAIN 101 155 KAZAL-LIKE.
FT DOMAIN 204 364 SERINE PROTEASE.
FT DOMAIN 365 467 PDZ.
FT ACT_SITE 220 220 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 250 250 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 328 328 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 480 AA; 51246 MW; 54BB9BA6C99A7BF4 CRC64;

HRA1_MOUSE Length: 480 January 30, 2004 15:08 Type: P Check: 7206 ..

1 MQSLRTLLS LLLLLLAAPS LALPSGTGRS APAATVCPEH CDPTRCAPPP

51 TDCGGRRVD ACQCEVCGA LEGAACGLQE GPCGEGLCV LPFGVPASAT

101 VRRRAQGLC VCASSPVCV SDAKYTNLC QLRAASRSE KLPQPVIVL

151 QRCAGCGQGE DPNSLRHKYN FIADVVEKFA PDVVKHELYR KLPFSKREVP

201 VASSGSGFVS EDGLIVTNAH VYTNKNRVK ELKNATVEA IIKVDVEKAD

```

```

251 IALIKIDHOG KLPVLLGRS SELRPGFEVY AIGSPFSLQN TTTTGIVSTT
301 QRGKELGLR NSDMYDIQTD AIINYGNSSG PLVNLGDEVI GINTLKVTAG
351 ISFAIPSDKI KKFLETSHDR QAKGKAVTKK KYIGIRMSL TSSKAKELKD
401 RHRDPDVLS GAYIEVIPD TPAEAGLKE NDVIISINGQ SVVTANDVSD
451 VIKKENTLNM VRRGNEDIV ITVIPEIDP

!!AA SEQUENCE 1.0
ID Q9KHD9 PRELIMINARY; PRT; 262 AA.
AC Q9KHD9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Enoyl-CoA hydratase-like protein.
OS Streptomyces griseus subsp. griseus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=67263;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM40695;
RX MEDLINE=20316791; PubMed=10858335;
RA Smith W.C., Xiang L., Shen B.;
RT "Genetic localization and molecular characterization of the nons gene
RT required for macrotetrolide biosynthesis in Streptomyces griseus
RT DSM40695."
RL Antimicrob. Agents Chemother. 44:1809-1817(2000).
DR EMBL; AF23011; AAF81231.1; -.
DR HSP; PI4604; 2DUB.
DR InterPro; IPR001753; EnCoA_hydrase.
DR Pfam; PF00378; ECH; 1.
DR PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
SQ SEQUENCE 262 AA; 27068 MW; D218FA2BDEC64BF2 CRC64;

Q9KHD9 Length: 262 January 30, 2004 15:08 Type: P Check: 1697 ..

1 MTPQETAPEG EPSPLLVVERH GRVATLTNR PHRRNAMSTA MLARLDHALG

51 KLTDQGEAP GALVLTGAGG TFSSGADTRE PDWRDLRRA VRAHFRITVF

101 AMLHRAFPV VAAVEGYALG GGLELALACD LVVAGEGALF GLPELGVNAV

151 PGCGAVHSIV RRACRGVAAR MLLIPGERVR ADELARLGAV ERTVPDGGAL

201 AEAQALAASV AAGDPALLEA GVLLRDSGH LORTAALGVE NGYWMQALSA

251 ANRDPSPSSG RP

!!AA SEQUENCE 1.0
ID Q8RUS6 PRELIMINARY; PRT; 107 AA.
AC Q8RUS6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE P0678F11.3 protein (P0413C03.30 protein).
GN P0678F11.3 OR P0413C03.30.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0678F11."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RL clone:P0413C03.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003437; BAB86096.1; --
DR EMBL; AP003451; BAB86154.1; --
DR Gramene; Q8RUS6; --
SQ SEQUENCE 107 AA; 11046 MW; 89097D882018D849 CRC64;

Q8RUS6 Length: 107 January 30, 2004 15:08 Type: P Check: 3612

1 MFSPPRAVNL RWSSLETEVE AAIAVEGCG VDLAWGRAL GLDPATVRLN
51 GTFVSRGRGH VSSAVTWRAL LDFFAARGLP TGDAPAAPVA VHKGPPPPPP
101 PPPVSVL

!!AA SEQUENCE 1.0
ID Q9QZK6 PRELIMINARY; PRT; 480 AA.
AC Q9QZK6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Insulin-like growth factor binding protein 5 protease.
GN PRS811.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Ovary;
RA Hourvitz A., Hennebold J.D., King G., Negishi H., Erickson G.F.,
RA Roby J.A., Mayo K.E., Adashi E.Y.;
RT "Mouse insulin-like growth factor binding protein 5-directed
RT endopeptidase: structural analysis, evolutionary analysis, ovarian
RT expression, hormonal regulation and cellular localization.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL; AF179369; AAD52682.1; --
DR MGD; MGI:1929076; Prs811.
DR InterPro; IPR000867; Insl_gro_fac_pr.
DR InterPro; IPR002350; kazal.
DR InterPro; IPR001478; PDZ.
DR STRAIN=C57BL/6; TISSUE=Ovary;
RA Hourvitz A., Hennebold J.D., King G., Negishi H., Erickson G.F.,
RA Roby J.A., Mayo K.E., Adashi E.Y.;
RT "Mouse insulin-like growth factor binding protein 5-directed
RT endopeptidase: structural analysis, evolutionary analysis, ovarian
RT expression, hormonal regulation and cellular localization.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL; AF179369; AAD52682.1; --
DR MGD; MGI:1929076; Prs811.
DR InterPro; IPR000867; Insl_gro_fac_pr.
DR InterPro; IPR002350; kazal.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00219; IGFBP; 1.
DR Pfam; PF00050; kazal; 1.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00121; IB; 1.
DR SMART; SM00280; KAZAL; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 480 AA; 51213 MW; 92BDDA85CF5B12B7 CRC64;

Q9QZK6 Length: 480 January 30, 2004 15:08 Type: P Check: 7594

1 MQSLRTLLS LLLLLLAAPS LALPSGTGRS APAATVCPEH CDPTRCAPP
51 TDCGGRVDR AGCCEVCGA LEGAACGLQE GPCGEGLOCV VPFGVPASAT
101 VRRRAQGLC VCASSEPCVG SDAKTYTNLC QLRAASRSE KLQPPVIVL
151 QRGACGQGE DPNLRHKYN FIADVVEKIA PAVVHIELYR KLPSKREVP
201 VASGSGFIVS EDGLIVTNAH VTNKNRVKV ELKNGATYEA KIKVDKAD
251 IALIKIDHQG KLPVLLGRS SELRPGEFVV AIGSPFSLQN TTTTGIVSTT

Q9QZK6 Length: 480 January 30, 2004 15:08 Type: P Check: 7594

```

```

301 QRGKGLGLR NSDMYIQT D AIINYNSGG PLVNLGDEVI GINTLKVTAG
351 ISFAIPSDKI KFLTQSHDR QAKGKAVTK KYIGIRMMSL TSSKAKELKD
401 RHRDFFPDVLS GAYIIEVPD TPAEAGGLKE NDIIVISNGQ SVVTANDVSD
451 VIKKENTLNM VVRGNEDIV ITVPEEIDP

!!AA SEQUENCE 1.0
ID Q91WS3 PRELIMINARY; PRT; 480 AA.
AC Q91WS3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Protease, serine, 11 (igf binding).
GN PRS811.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL; BC013516; AAH13516.1; --
DR MGD; MGI:1929076; Prs811.
DR InterPro; IPR000867; Insl_gro_fac_pr.
DR InterPro; IPR002350; kazal.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00219; IGFBP; 1.
DR Pfam; PF00050; kazal; 1.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00121; IB; 1.
DR SMART; SM00280; KAZAL; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 480 AA; 51212 MW; 76BDD5E862EDC9DA CRC64;

Q91WS3 Length: 480 January 30, 2004 15:08 Type: P Check: 7882

1 MQSLRTLLS LLLLLLAAPS LALPSGTGRS APAATVCPEH CDPTRCAPP
51 TDCGGRVDR AGCCEVCGA LEGAACGLQE GPCGEGLOCV VPFGVPASAT
101 VRRRAQGLC VCASSEPCVG SDAKTYTNLC QLRAASRSE KLQPPVIVL
151 QRGACGQGE DPNLRHKYN FIADVVEKIA PAVVHIELYR KLPSKREVP
201 VASGSGFIVS EDGLIVTNAH VTNKNRVKV ELKNGATYEA KIKVDKAD
251 IALIKIDHQG KLPVLLGRS SELRPGEFVV AIGSPFSLQN TTTTGIVSTT
301 QRGKGLGLR NSDMYIQT D AIINYNSGG PLVNLGDEVI GINTLKVTAG
351 ISFAIPSDKI KFLTQSHDR QAKGKAVTK KYIGIRMMSL TSSKAKELKD
401 RHRDFFPDVLS GAYIIEVPD TPAEAGGLKE NDIIVISNGQ SVVTANDVSD
451 VIKKENTLNM VVRGNEDIV ITVPEEIDP

!!AA SEQUENCE 1.0
ID Q9QZK5 PRELIMINARY; PRT; 480 AA.
AC Q9QZK5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Insulin-like growth factor binding protein 5 protease.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Ovary;
RA Hourvitz A., Hennebold J.D., King G., Negishi H., Erickson G.F.,
RA Roby J.A., Mayo K.E., Adashi E.Y.;
RT "Mouse insulin-like growth factor binding protein 5-directed
RT endopeptidase: structural assessment, evolutionary analysis, ovarian
RT expression, hormonal regulation and cellular localization.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL; AF179370; AAD52683.1; -.
DR MEROPS; S01.277; -.
DR InterPro; IPR000867; Insl_gro_fac_pr.
DR InterPro; IPR002350; kazal.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00219; IGFBP; 1.
DR Pfam; PF00050; kazal; 1.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00121; IB; 1.
DR SMART; SM00280; KAZAL; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 480 AA; 51330 MW; 37A864C5A8FFC035 CRC64;
Q9QZK5 Length: 480 January 30, 2004 15:08 Type: P Check: 8689 ..
1 MQFLRTALLS LLLLLLAAPS LALPSGISRS APAATVCPEH CDPTRCAPPP
51 TDCEGGRVRD AGCCCEVCGA LEGAVCGLOE GPCGEGLOCV VPFGVPASAT
101 VRRRAQAGLC VCASSEPCVG SDAKTYTNLC QLRAASRRSE KLRQPPVIVL
151 QRGACGGQGE DPNSLRHKYN FIADVVEKIA PAVVHIELYR KLPFSKREVP
201 VAGSGGFIVS EDGLIVTNAH VVTNKNRVKV ELKNGATYEA KIKQDVDEKAD
251 IALIKIDHQG KLPVLLGRS SELRPGFVW AIGSPFSLQN TVTTGIVSTT
301 QRGKELGLR NSDMYIQTD AIINFGNSGG PLVNLGDEVI GINTLKVTAG
351 ISFAIPSDKI KKFLTESHDR QAKGKVTYKK KYIGIRMMSL TSSKAKELKD
401 RHRDFPDVIS GAVIEVIPD TPAEAGGLKE NDVIISNGQ SVVTANDVSD
451 VIKKENTLNM VVRGNEDIV ITVPPEIDP

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 00:13:12 ; Search time 8.97183 Seconds
(without alignments)
225.098 Million cell updates/sec

Title: US-09-461-684C-1
Perfect score: 109
Sequence: 1 CKKKKKKKKKKKKKKKKKK 21
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	100	91.7	215	2	I52523
2	100	91.7	380	2	T46395
3	100	91.7	517	2	T49173
4	93	85.3	166	2	T18513
5	90	82.6	483	2	F71619
6	87	79.8	107	2	C86477
7	84	77.1	441	2	A48455
8	81	74.3	4550	2	T18440
9	80	73.4	784	2	T18452
10	78	71.6	529	2	T50609
11	78	71.6	560	2	T06377
12	78	71.6	1560	2	T42727
13	77	70.6	229	2	JC7219
14	77	70.6	266	2	A86288
15	77	70.6	3724	2	T18427
16	76	69.7	511	2	S58322
17	76	69.7	812	2	S43604
18	75	68.8	474	2	T38485
19	75	68.8	504	2	S48550
20	74	67.9	191	2	T23594
21	74	67.9	198	2	D97301
22	74	67.9	414	2	F64386
23	74	67.9	483	2	S41853
24	73	67.0	212	2	T49559
25	72.5	66.5	683	2	T00872
26	72	66.1	404	2	A75192
27	72	66.1	560	2	T29586
28	72	66.1	678	2	A54514
29	72	66.1	1701	2	T09127

30	72	66.1	4981	2	T18489	hypothetical prote
31	71	65.1	683	2	T34103	hypothetical prote
32	71	65.1	686	1	A44842	cGMP-gated ion cha
33	71	65.1	690	2	A42161	cGMP-gated cation
34	70.5	64.7	163	2	T42696	hypothetical prote
35	70	64.2	222	2	T28919	hypothetical prote
36	70	64.2	233	2	S55165	hypothetical prote
37	70	64.2	291	2	T48617	hypothetical prote
38	70	64.2	451	2	T16418	hypothetical prote
39	70	64.2	497	2	T29814	hypothetical prote
40	70	64.2	508	2	E71620	hypothetical prote
41	70	64.2	688	2	B42161	cGMP-gated cation
42	70	64.2	2013	2	C71610	probable membrane
43	69.5	63.8	2231	2	S53416	SEN1 protein - yea
44	69	63.3	142	2	S54481	hypothetical prote
45	69	63.3	167	2	S38112	hypothetical prote

ALIGNMENTS

RESULT 1

I52523
nucleoporin p62 homolog - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 05-Nov-1999
C:Accession: I52523
R:Wang, Z.Q.; Akmal, K.M.; Kim, K.H.
Biol. Reprod. 51, 1022-1030, 1994
A:Title: An unusual nucleoporin-related messenger ribonucleic acid is present in the germ
A:Reference number: I52523; MUID:95151924; PMID:7849178
A:Accession: I52523
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-215 <RES>
A:Cross-references: GB:S75997; NID:G913245; PIDN:AAB33384.1; PID:G913246
A:Experimental source: testis

Query Match 91.7%; Score 100; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	2	KKKKKKKKKKKKKKKKKKK	21
Db	35	KKKKKKKKKKKKKKKKKKK	54

RESULT 2

T46395
hypothetical protein DKFZp434I1120.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46395
R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23031
A:Accession: T46395
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-380 <AAA>
A:Cross-references: EMBL:AL137556
A:Experimental source: adult testis; clone DKFZp434I1120
C:Genetics:
A:Note: DKFZp434I1120.1

Query Match 91.7%; Score 100; DB 2; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	2	KKKKKKKKKKKKKKKKKKK	21
Db	355	KKKKKKKKKKKKKKKKKKK	374

RESULT 3

T49173
 hypothetical protein T20N10.250 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
 C:Accession: T49173
 R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z25017
 A:Accession: T49173
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-517 <DAN>
 A:Cross-references: ENBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250
 A:Experimental source: cultivar Columbia; BAC clone T20N10
 C:Genetics:

Query Match 91.7%; Score 100; DB 2; Length 517;
 Best Local Similarity 100.0%; Pred. No. 0.003;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 KKKKKKKKKKKKKKKKKKK 21
 Db 444 KKKKKKKKKKKKKKKKKKK 463

C:Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

RESULT 4

T18513
 hypothetical protein C0845c - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T18513

R:Lawson, D.; Bowman, S.; Barrell, B.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18513

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-166 <LAW>
 A:Cross-references: ENBL:Z98551; PIDN:CABL1123.2
 C:Genetics:
 A:Map position: 3
 A:Introns: 19/1
 A:Note: C0845c

Query Match 85.3%; Score 93; DB 2; Length 166;
 Best Local Similarity 90.0%; Pred. No. 0.0067;
 Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKK 21
 Db 37 KKKKKKKKKKKKKKKKKKK 56

RESULT 5

F71619
 hypothetical protein PFB0235w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: F71619

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ; Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551
 A:Accession: F71619
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-483 <GAR>

A:Cross-references: GB:AE001382; GB:AB001362; NID:G3845130; PIDN:AAF79354.1; GSPDB:GN00141
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0235w

Query Match 92.6%; Score 90; DB 2; Length 483;
 Best Local Similarity 85.0%; Pred. No. 0.026;
 Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKK 21
 Db 449 KKKKKKKKKKKKKKKKKKK 468

RESULT 6

C86477
 protein F1504.29 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C86477

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 anson, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: C86477

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-107 <SPO>

A:Cross-references: GB:AE005172; NID:G8778346; PIDN:AAF79354.1; GSPDB:GN00141

C:Genetics:

A:Gene: F1504.29

A:Map position: 1

Query Match 79.8%; Score 87; DB 2; Length 107;
 Best Local Similarity 94.4%; Pred. No. 0.019;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKK 19
 Db 29 KKKKKKKKKKKKKKKKKKK 46

RESULT 7

A48455
 acidic phosphoprotein PCEMA1q - Plasmodium chabaudi

C:Species: Plasmodium chabaudi

C>Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998

C:Accession: A48455

R:Deleersnijder, W.; Prasomsitti, P.; Tungpradubkul, S.; Hendrix, D.; Hamers-Casterman, C.

Mol. Biochem. Parasitol. 56, 59-68, 1992

A:Title: Structure of a Plasmodium chabaudi acidic phosphoprotein that is associated with

A:Reference number: A48455; MUID:93116806; PMID:1475002

A:Accession: A48455

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-441

A:Cross-references: GB:M95789; NID:G160602; PID:G160603

A:Experimental source: IP-PC1/C

A:Note: sequence extracted from NCBI backbone (NCBI:121415, NCBI:121416)

C:Keywords: phosphoprotein

Query Match 77.1%; Score 84; DB 2; Length 441;
 Best Local Similarity 80.0%; Pred. No. 0.089;
 Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||:
Db 394 KKKKKKKKKKKKKKKKKKKKK 413

RESULT 8

T18440
hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18440
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18440
A:Molecule type: DNA
A:Residues: 1-4550 <LAW>
A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAB11121.1
C:Genetics:
A:Map position: 3
A:Note: C0425w

Query Match 74.3%; Score 81; DB 2; Length 4550;
Best Local Similarity 44.2%; Pred. No. 0.77;
Matches 19; Conservative 2; Mismatches 0; Indels 22; Gaps 1;
QY 1 CKKK-----KKKKKKKKKKKKKK 21
|||||:|||||:|||||:
Db 132 CKKKYFNIKNKYNKPYQNNIKKKKKKKKKKKKK 174

RESULT 9

T18452
hypothetical protein C0560c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18452
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z18937
A:Accession: T18452
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-784 <LAW>
A:Cross-references: EMBL:AL008970; NID:e1407852; PID:e1332545; PIDN:CAA15594.1
C:Genetics:
A:Map position: 3
A:Note: C0560c

Query Match 73.4%; Score 80; DB 2; Length 784;
Best Local Similarity 75.0%; Pred. No. 0.31;
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||:
Db 360 EKXKQKKKKKKKKKKKKKKKK 379

RESULT 10

T50609
hypothetical protein DKFZp761B2423.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50609
R:Blocker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, June 2000
A:Reference number: Z25143
A:Accession: T50609
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-529 <AAA>
A:Cross-references: EMBL:AL359564

A:Experimental source: adult amygdala; clone DKFZp761B2423
C:Genetics:
A:Note: DKFZp761B2423.1

Query Match 71.6%; Score 78; DB 2; Length 529;
Best Local Similarity 80.0%; Pred. No. 0.37;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||:
Db 464 KKKKKKKKKKKKKKKKKKKKK 483

RESULT 11

T06377
SAR DNA-binding protein-1 - garden pea
C:Species: Pisum sativum (garden pea)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 02-Jun-2000
C:Accession: T06377
R:Hatton, D.; Gray, J.C.
submitted to the EMBL Data Library, April 1998
A:Description: cDNA encoding a pea SAR DNA-binding protein that shows homology to nucleol
A:Reference number: Z15637
A:Accession: T06377
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-560 <HAT>
A:Cross-references: EMBL:AF061962; NID:g3132695; PIDN:AAC16330.1; PID:g3132696
C:Genetics:
A:Gene: SARBP-1
C:Superfamily: garden pea SAR DNA-binding protein

Query Match 71.6%; Score 78; DB 2; Length 560;
Best Local Similarity 75.0%; Pred. No. 0.38;
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||:
Db 463 KKKKKKKKKKKKKKKKKKKKK 482

RESULT 12

T42727
proliferation potential-related protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000
C:Accession: T42727
R:Witte, M.W.; Scott, R.E.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z22246
A:Accession: T42727
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1560 <WIT>
A:Cross-references: EMBL:U89913; NID:g3858884; PID:g3858885; PIDN:AAC72432.1
A:Experimental source: strain Balb/c
C:Genetics:
A:Gene: P2P-R
C:Function:
A:Description: involved in hnRNP association and Rb1 binding
A:Superfamily: RING finger homology
F:57-107/Domain: RING finger homology <RRN>

Query Match 71.6%; Score 78; DB 2; Length 1560;
Best Local Similarity 80.0%; Pred. No. 0.74;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||:
Db 1497 KKKKKKKKKKKKKKKKKKKKK 1516

RESULT 13

JC7219
 nuclear protein SR-25 - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
 C:Accession: JC7219
 R:Sasahara, K.; Yanoaka, T.; Moritani, M.; Tanaka, M.; Iwahana, H.; Yoshimoto, K.; Miyag
 Biochem. Biophys. Res. Commun. 269, 444-450, 2000
 A:Title: Molecular cloning and expression analysis of a putative nuclear protein, SR-25.
 A:Reference number: JC7219; MUID:20175222; PMID:10708573
 A:Accession: JC7219
 A:Molecule type: mRNA
 A:Residues: 1-229 <SAS>
 A:Cross-references: DBJ:AB035383; NID:g7619895; PIDN:BAA94743.1; PID:g7619896
 A:Experimental source: MIN6 cell line
 C:Comment: This protein is a highly hydrophilic nuclear protein with a serine-arginine x
 A splicing factors.
 C:Keywords: nucleus; RNA processing

Query Match 70.6%; Score 77; DB 2; Length 229;
 Best Local Similarity 75.0%; Pred. No. 0.27;
 Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
 Db 95 KKKKKKKKKKKKKKKKKKKKK 114

RESULT 14
 A86288
 protein F9L1.30 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: A86288
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiz, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: A86288
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-266 <STO>
 A:Cross-references: GB:AE005172; NID:gs103832; PIDN:AAD39662.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F9L1.30
 A:Map position: 1

Query Match 70.6%; Score 77; DB 2; Length 266;
 Best Local Similarity 70.0%; Pred. No. 0.3;
 Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
 Db 231 KKKKKKKKKKKKKKKKKKKKK 250

RESULT 15
 T18427
 hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18427
 R:Lawson, D.; Bowman, S.; Barrell, B.
 Submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18427
 A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-3724 <LAW>
 A:Cross-references: EMBL:Z98547; NID:el325376; PID:el325379; PIDN:CAB111104.1
 C:Genetics:
 A:Introns: 307/1; 1545/2
 A>Note: C0335C

Query Match 70.6%; Score 77; DB 2; Length 3724;
 Best Local Similarity 75.0%; Pred. No. 1.6;
 Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 CXXXXXXXXXXXXXXXXXXXXX 20
 Db 2201 CEIKKKKKKKKKKKKKKKKKKK 2220

Search completed: January 30, 2004, 00:26:20
 Job time : 8.97183 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:27 ; Search time 4.73239 Seconds
(without alignments)
208.681 Million cell updates/sec

Title: US-09-461-684C-1

Perfect score: 109

Sequence: 1 CKKKKKKKKKKKKKKKKK 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	77.1	441	1 PHPA_PLACH	Q02752 plasmodium
2	76	69.7	511	1 NOP5_YEAST	Q12499 saccharomyc
3	75	68.8	474	1 CBP5_SCHPO	Q14007 schizosacch
4	75	68.8	504	1 SKL1_YEAST	Q12460 saccharomyc
5	75	68.8	2404	1 SON_MOUSE	Q9qx47 mus musculu
6	75	68.8	2426	1 SON_HUMAN	P18583 homo sapien
7	74	67.9	414	1 Y694_METJA	Q58105 methanococ
8	74	67.9	474	1 CBP5_KLJLA	Q13473 kluyveromyc
9	74	67.9	483	1 CBP5_YEAST	P23322 saccharomyc
10	74	67.9	726	1 BRD3_HUMAN	Q15059 homo sapien
11	72	66.1	479	1 CBP5_CANAL	Q43101 candida alb
12	72	66.1	678	1 GARP_PLAFF	P13816 plasmodium
13	71	65.1	351	1 CG79_HUMAN	Q9y388 homo sapien
14	71	65.1	686	1 CNG1_HUMAN	P29973 h cgm-p-gate
15	70.5	64.7	1411	1 TCOF_HUMAN	Q13428 homo sapien
16	70	64.2	233	1 YJ08_YEAST	P47006 saccharomyc
17	70	64.2	599	1 HNZ1_HUMAN	Q9ugus homo sapien
18	70	64.2	683	1 CNG1_RAT	Q62927 r cgm-p-gate
19	70	64.2	684	1 CNG1_MOUSE	P29974 m cgm-p-gate
20	69.5	63.8	534	1 NOP5_RAT	Q9qz86 rattus norv
21	69.5	63.8	2231	1 SEN1_YEAST	Q00416 saccharomyc
22	69	63.3	142	1 YN8H_YEAST	Q03525 saccharomyc
23	69	63.3	167	1 YK20_YEAST	P36133 saccharomyc
24	69	63.3	723	1 BLVR_DROME	Q05344 drosophila
25	69	63.3	843	1 BLVR_BOVIN	Q03368 bos taurus
26	68.5	62.8	724	1 Y061_CABEL	P24600 caenorhabdi
27	68	62.4	523	1 DRP3_YEAST	P20447 saccharomyc
28	68	62.4	1178	1 MNV4_YEAST	P36044 saccharomyc
29	67	61.5	118	1 Y093_CABEL	Q00032 caenorhabdi
30	67	61.5	690	1 CNG1_BOVIN	Q00194 b cgm-p-gate
31	67	61.5	691	1 CNG1_CANFA	Q28279 c cgm-p-gate
32	67	61.5	1002	1 IF2P_YEAST	P39730 saccharomyc
33	67	61.5	1220	1 IF2P_HUMAN	O60841 homo sapien

34	67	61.5	1362	1 BRD4_HUMAN	O60885 homo sapien
35	67	61.5	2280	1 YCF2_OENHO	Q9mef2 oenothera h
36	66	60.6	481	1 CBP5_EMENI	Q43100 emeritella
37	66	60.6	487	1 CBP5_ASPFU	O43102 aspergillus
38	66	60.6	667	1 KYV1_SCHPO	O13796 schizosacch
39	66	60.6	1153	1 A3D1_HUMAN	O14617 homo sapien
40	66	60.6	1240	1 YNJ1_YEAST	P53935 saccharomyc
41	65.5	60.1	508	1 NO60_DROME	O44081 drosophila
42	65.5	60.1	514	1 DKC1_HUMAN	O60832 homo sapien
43	65	59.6	217	1 KSI_HYDAT	P38978 hydra atten
44	65	59.6	271	1 YG5W_YEAST	P53335 saccharomyc
45	65	59.6	320	1 YD33_YEAST	Q12117 saccharomyc

ALIGNMENTS

RESULT 1
ID PHPA_PLACH STANDARD; PRT; 441 AA.
AC Q02752;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Acidic phosphoprotein precursor (50 kDa antigen).
GN PCEMAL.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IP-PCI;
RX MEDLINE=93116806; PubMed=1475002;
RA Deleersnijder W., Pracomitti P., Tungpradubkul S., Hendrix D.,
RA Hamers-Casterman C., Hamers R.;
RT "Structure of a Plasmodium chabaudi acidic phosphoprotein that is
associated with the host erythrocyte membrane.";
RL Mol. Biochem. Parasitol. 56:59-68(1992).
CC -!- FUNCTION: DURING INFECTION, THIS PHOSPHOPROTEIN PROBABLY MODULATES
THE STRUCTURE OF THE RED CELL MEMBRANE TO THE ADVANTAGE OF THE
PARASITE, ALTHOUGH ITS PRECISE FUNCTION IS NOT KNOWN.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ON THE
CYTOPLASMIC FACE OF THE HOST ERYTHROCYTE MEMBRANE.
CC -!- MISCELLANEOUS: ASSOCIATED WITH THE HOST RED CELL MEMBRANE
THROUGHOUT THE ENTIRE ERYTHROCYTIC CYCLE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: M95789; AAA29732.1; --
CC PIR: A48455; A48455.
DR Phosphorylation; Signal; Antigen; Membrane; Repeat; Erythrocyte.
KW SIGNAL 1 15 OR 24 (POTENTIAL).
FT CHAIN 16 441 ACIDIC PHOSPHOPROTEIN.
FT DOMAIN 186 313 16 X 8 AA TANDEM REPEATS.
FT REPEAT 186 193 1-1.
FT REPEAT 194 201 1-2.
FT REPEAT 202 209 1-3.
FT REPEAT 210 217 1-4.
FT REPEAT 218 225 1-5.
FT REPEAT 226 233 1-6.
FT REPEAT 234 241 1-7.
FT REPEAT 242 249 1-8.
FT REPEAT 250 257 1-9.
FT REPEAT 258 265 1-10.
FT REPEAT 266 273 1-11.
FT REPEAT 274 281 1-12.
FT REPEAT 282 289 1-13.

FT REPEAT 290 297 1-14.
 FT REPEAT 298 305 1-15.
 FT REPEAT 306 313 1-16.
 FT DOMAIN 353 370 2 X 9 AA TANDEM REPEATS.
 FT REPEAT 353 360 2-1.
 FT REPEAT 361 368 2-2.
 FT DOMAIN 371 417 LYS-RICH (BASIC).
 FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 441 AA; 49708 MW; DB85E83E795EE7B5 CRC64;

Query Match 77.1%; Score 84; DB 1; Length 441;
 Best Local Similarity 80.0%; Pred. No. 0.032;
 Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
 DB 394 KKKKKKKKKKKKKKKKKKK 413

RESULT 2

NOPS_YEAST

ID NOPS_YEAST STANDARD; PRT; 511 AA.

AC Q12499;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nucleolar protein NOP58 (Nucleolar protein NOPS).
 GN NOP58 OR NOPS OR YOR310C OR O6108.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / FY1679;
 RA Pearson B.M., Hernando Y., Wolf S.S., Kalogeropoulos A., Schweizer M.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=W303;
 RX MEDLINE=98298165; PubMed=9632712;
 RA Wu P., Brockenbrough J.S., Metcalfe A.C., Chen S., Aris J.P.;
 RT "Nop58 is a small nucleolar ribonucleoprotein component required for
 pre-18S rRNA processing in yeast.";
 RL J. Biol. Chem. 273:16453-16463 (1998).
 CC -!- FUNCTION: REQUIRED FOR PRE-18S RNA PROCESSING. MAY BIND
 MICROTUBULES.
 CC -!- SUBUNIT: INTERACTS WITH NOP56 AND NOP1.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE NOPS/NOP56 FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X90565; CAA62165.1; -;
 DR EMBL; 275217; CAA99630.1; -;
 DR EMBL; AF056070; AAC39484.1; -;
 DR PIR; S58322; S58322.
 DR SGD; S0005837; NOP58.
 DR GO; GO:0005732; C:small nucleolar ribonucleoprotein complex; IPI.
 DR GO; GO:0003754; F:chaperone activity; NAS.
 DR GO; GO:0017069; F:sRNA binding activity; IDA.
 DR GO; GO:0030490; P:processing of 20S pre-rRNA; IPI.
 DR GO; GO:0006608; P:sRNP protein-nucleus import; NAS.
 DR InterPro; IPR002687; Nop.
 DR Pfam; PF01798; Nop; 1.
 DR ProDom; PD004104; Nop; 1.

KW Ribosome biogenesis; Nuclear protein; rRNA processing.
 FT DOMAIN 441 511 ASP/GLU/LYS-RICH.
 SQ SEQUENCE 511 AA; 56956 MW; 8A289448B2A19E2 CRC64;

Query Match 69.7%; Score 76; DB 1; Length 511;
 Best Local Similarity 70.0%; Pred. No. 0.21;
 Matches 14; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKKKK 21
 DB 480 KKKKKKKKKKKKKKKKKKK 499

RESULT 3

CBF5_SCHPO

ID CBF5_SCHPO STANDARD; PRT; 474 AA.

AC O14007;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Centromere/microtubule binding protein cbf5 (Centromere-binding factor
 5) (Nucleolar protein cbf5).
 GN CBF5 OR SPAC29A4.04C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voiclaert G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880 (2002).
 CC -!- FUNCTION: BINDS DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
 CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
 CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
 CC -!- SIMILARITY: Contains 1 PUA domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL; Z97210; CAB10131.1; -.
DR PIR; T38485; T38485.
DR GenBank; SPombe; SPAC29A4.04c; -.
DR InterPro; IPR004802; Cbf5.
DR InterPro; IPR002478; PUA.
DR InterPro; IPR002501; TruB_N.
DR InterPro; IPR004521; Unchar_dom_2.
DR Pfam; PF01472; PUA; 1.
DR Pfam; PF01509; TruB_N; 1.
DR SMART; SM00359; PUA; 1.
DR TIGRFAMs; TIGR00425; Cbf5; 1.
DR TIGRFAMs; TIGR00451; unchar_dom_2; 1.
DR PROSITE; PS50890; PUA; 1. Repeat; Nuclear protein; DNA-binding.
KW Microtubules; Centromere; Repeat.
FT DOMAIN 271 346
FT DOMAIN 434 468
FT REPEAT 443 445
FT REPEAT 450 452
FT REPEAT 454 456
FT REPEAT 457 459
FT REPEAT 460 462
FT REPEAT 463 465
FT REPEAT 466 468
FT REPEAT 474 476
FT SEQUENCE 474 AA; 53110 MW; B8C9896C5FAEBB71 CRC64;
SQ
Query Match 68.8%; Score 75; DB 1; Length 474;
Best Local Similarity 73.1%; Pred. No. 0.24;
Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 20
DB 454 KKKKKKKKKKKKKKKKK 472

RESULT 4
SIK1_YEAST STANDARD; PRT; 504 AA.
AC Q12460;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE SIK1 protein (Nucleolar protein NOP56).
GN SIK1 OR NOP56 OR YLR197W OR L8167.9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9288C / YPH1;
RX MEDLINE=96040178; PubMed=7547500;
RA Morin P.J., Downs J.A., Snodgrass A.M., Gilmore T.D.;
RT "Genetic analysis of growth inhibition by GAL4-L kappa B-alpha in
RT Saccharomyces cerevisiae."
RN Cell Growth Differ. 6:789-798 (1995).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=9288C / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansoorge W.,
RA Benes V., Brueckner M., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Fortelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharif M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA Vierdeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
```

```

RL Nature 387:87-90 (1997).
RN [3]
RP CHARACTERIZATION, AND MUTAGENESIS.
RX MEDLINE=98038777; PubMed=9372940;
RA Gautier T., Berges T., Tollervy D., Hurt E.;
RT "Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Noplp
RT and are required for ribosome biogenesis."
RL Mol. Cell. Biol. 17:7088-7098 (1997).
CC -1- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT SYNTHESIS.
CC -1- SUBUNIT: INTERACTS WITH NOP1 AND NOP58.
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- SIMILARITY: BELONGS TO THE NOPS/NOP56 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U20237; AAC49066.1; -.
DR EMBL; U14913; AAB67431.1; -.
DR PIR; S48550; S48550.
DR SGD; S0004187; SIK1.
DR GO; GO:0005732; C:small nucleolar ribonucleoprotein complex; IPI.
DR GO; GO:0030490; P:processing of 20S pre-rRNA; IPI.
DR InterPro; IPR002687; Nop.
DR Pfam; PF01798; Nop; 1.
DR ProDom; PD004104; Nop; 1.
KW Ribosome biogenesis; Nuclear protein.
FT DOMAIN 443 504
FT MUTAGEN 333 333
FT MUTAGEN 355 355
FT MUTAGEN 385 385
FT MUTAGEN 504 AA; 56864 MW; F8522A5870EF4842 CRC64;
SQ
Query Match 68.8%; Score 75; DB 1; Length 504;
Best Local Similarity 70.0%; Pred. No. 0.25;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 465 KKKKKKKKKKKKKKKKK 484

RESULT 5
SON_MOUSE STANDARD; PRT; 2404 AA.
AC Q9QX47; Q9QX47; Q9QX47; Q9QX47; Q9QX47;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE SON protein.
GN SON.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=129/Sv.
RX MEDLINE=20408986; PubMed=10950926;
RA Wynn S.D., Fisher R.A., Page C., Price M., Liu Q.Y., Khan I.M.,
RA Zammit P., Dadrah K., Mazzari W., Kessling A., Lee J.S., Buluwa L.;
RT "Organization and conservation of the GART/SON/DONSON locus in mouse
RT and human genomes."
RL Genomics 68:57-62 (2000).
RN [2]
```

RP SEQUENCE OF 1-116 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus, Small intestine, and Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa K., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kaeukawa T., Saito R.,
RA Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Glessi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuoka Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gusincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690 (2001).
CC -!- FUNCTION: Transcriptional repressor. Binds to the consensus DNA
CC sequence: 5'-GA[GT]AN[CG]AG[CC-3'. Might protect cells from
CC apoptosis. Might be involved in pre-mRNA splicing (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9QX47-1; Sequences=Displayed;
CC Name=2;
CC IsoId=Q9QX47-2; Sequences=VSP_004416, VSP_004417;
CC -!- TISSUE SPECIFICITY: Widely expressed.
CC -!- DOMAIN: Contains 8 types of repeats which are distributed in 3
CC regions.
CC -!- SIMILARITY: Contains 1 G-patch domain.
CC -!- SIMILARITY: Contains 1 DREM (double-stranded RNA-binding) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF193606; AAF23120.1; -
DR EMBL; AF193595; AAF23120.1; JOINED.
DR EMBL; AF193596; AAF23120.1; JOINED.
DR EMBL; AF193597; AAF23120.1; JOINED.
DR EMBL; AF193598; AAF23120.1; JOINED.
DR EMBL; AF193599; AAF23120.1; JOINED.
DR EMBL; AF193600; AAF23120.1; JOINED.
DR EMBL; AF193601; AAF23120.1; JOINED.
DR EMBL; AF193602; AAF23120.1; JOINED.
DR EMBL; AF193603; AAF23120.1; JOINED.
DR EMBL; AF193604; AAF23120.1; JOINED.
DR EMBL; AF193605; AAF23120.1; JOINED.
DR EMBL; AF193607; AAF23120.1; -
DR EMBL; AK019312; BAB31659.1; -
DR EMBL; AK019081; BAB31536.1; -
DR EMBL; AK008478; BAB25691.1; -
DR EMBL; AK008256; BAB25562.1; -
DR MGD; MGI:98353; Son.
DR GO; GO:0005515; F:protein binding activity; IPI.
DR InterPro; IPR001159; DS_RBD.
DR InterPro; IPR000467; G_patch.
DR Pfam; PF00035; dsrm; 1.
DR Pfam; PF01585; G_patch; 1.
DR SMART; SM00443; G_patch; 1.
DR PROSITE; PS50137; DS_RBD; 1.
DR PROSITE; PS50174; G_PATCH; 1.

KW RNA-binding; DNA-binding; Nuclear protein; Repeat;
KW Alterative splicing
FT DOMAIN 721 850 13 X 10 AA TANDEM REPEATS OF L-A-[ST]-
FT [NSG]-[TS]-MDSQM.
FT 11 X 7 AA TANDEM REPEATS OF [DR]-P-Y-R-
FT [LI] [AG] [QHP].
FT 14 X 6 AA REPEATS OF [ED]-R-S-M-M-S.
FT 1-1.
FT REPEAT 961 966
FT REPEAT 969 974
FT REPEAT 976 981
FT REPEAT 985 990
FT REPEAT 993 998
FT REPEAT 1001 1006
FT REPEAT 1010 1015
FT REPEAT 1018 1023
FT REPEAT 1026 1031
FT REPEAT 1035 1040
FT REPEAT 1044 1049
FT REPEAT 1055 1060
FT REPEAT 1066 1071
FT REPEAT 1075 1080
FT DOMAIN 1101 1133 3 X 11 AA TANDEM REPEATS OF P-P-L-P-P-E-E-
FT P-P-[TME]-[MTG].
FT 7 X 7 AA REPEATS OF P-S-R-R-S-R-[TS].
FT 2-1.
FT REPEAT 1310 1379
FT REPEAT 1310 1316
FT REPEAT 1338 1344
FT REPEAT 1345 1351
FT REPEAT 1352 1358
FT REPEAT 1359 1365
FT REPEAT 1366 1372
FT REPEAT 1373 1379
FT DOMAIN 1319 1390 2 X 19 AA REPEATS OF P-S-R-R-R-R-S-R-S-V-
FT V-R-R-S-S-F-S-I-S.
FT REPEAT 1319 1337 3-1.
FT REPEAT 1380 1390 3-2 (APPROXIMATE).
FT DOMAIN 1391 2017 3 X TANDEM REPEATS OF [ST]-P-[VLI]-R-
FT [RL]-[RK]-[RF]-S-R.
FT G-PATCH.
FT DOMAIN 2283 2329
FT DOMAIN 2349 2404
FT VARSPLIC 2086 2086 /FTid=VSP_004416.
FT VARSPLIC 2087 2404 Missing (in isoform 2).
FT VARSPLIC 2087 2404 /FTid=VSP_004417.
SQ SEQUENCE 2404 AA; 261428 MW; 648BF28ED3FC01D9 CRC64;
Query Match 58.8%; Score 75; DB 1; Length 2404;
Best Local Similarity 75.0%; Pred. No. 0.85;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 2 KKKKKKKKKKKKKKKKKKK 21
Db 109 KKKKKKKKKKKKKKKKKKK 128
RESULT 6
SON_HUMAN
ID SON_HUMAN STANDARD; PRT; 2426 AA.
AC F18583; Q14487; Q95981; Q14120; Q9H7B1; Q9P070; Q9P072; Q9UKP9;
AC Q9UPY0;
DT 01-NOV-1990 (Rel. 16, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE SON protein (SON3) (Negative regulatory element-binding protein) (NRE-
DE binding protein) (DBP-5) (bax antagonist selected in saccharomyces 1)
DE (BASSI) (protein C21orf50).
GN SON OR NREBP OR DBP5 OR C21ORF50 OR KIAA1019.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B; C; D; E AND F).
RX MEDLINE=21564202; PubMed=11707072;

RA Raymond A., Friedli M., Neergaard Henriksen C., Chapot F.,
RA Deutscher S., Ucla C., Rossier C., Lyle R., Guipponi M.,
RA Antonarakis S.E.;
RT "From PRDs and open reading frames to cDNA isolation: revisiting the
RT human chromosome 21 transcription map";
RL Genomics 78:46-54(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM G).
RC TISSUE=Liver;
RX MEDLINE=21316479; PubMed=11306577;
RA Sun C.-T., Lo W.-Y., Wang I.-H., Lo Y.-H., Shiou S.-R., Lai C.-K.,
RA Ting L.-P.;
RT "Transcription repression of human hepatitis B virus genes by negative
RT regulatory element-binding protein/SON";
RL J. Biol. Chem. 276:24059-24067(2001).
RN [3]
RP SEQUENCE OF 1-689 FROM N.A. (ISOFORM H).
RC TISSUE=Placenta;
RA Casadei R., Strippoli P., D'Addabbo P., Canaider S., Lenzi L.,
RA Vitale L., Giannone S., Carinci P., Zannotti M.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-130 FROM N.A.
RC TISSUE=Smooth muscle;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA Odayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isoqai T., Sugano S.;
RT "NEDO human cDNA sequencing project";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-114 FROM N.A.
RC TISSUE=Blood;
RA Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS from cd34+ stem cells";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 437-2426 FROM N.A. (ISOFORM B).
RC TISSUE=Brain;
RX MEDLINE=9397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hiroseawa M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro";
RL DNA Res. 6:197-205(1999).
RN [7]
RP SEQUENCE OF 554-2426 FROM N.A. (ISOFORM A).
RX MEDLINE=92049296; PubMed=1944255;
RA Chumakov I.M., Berdichevskii F.B., Sokolova N.V., Reznikov M.V.,
RA Prasolov V.S.;
RT "Identification of a protein product of a novel human gene SON and
RT the biological effect upon administering a changed form of this gene
RT into mammalian cells";
RL Mol. Biol. (Mosk) 25:731-740(1991).
RN [8]
RP SEQUENCE OF 709-1079 FROM N.A. (ISOFORM I).
RC TISSUE=Placenta;
RX MEDLINE=93062885; PubMed=1435774;
RA Bliskovskii V.V., Kirillov A.V., Zakhariev V.M., Chumakov I.M.;
RT "The human son gene: the large and small transcripts contains various
RT 5'-terminal sequences";
RL Mol. Biol. (Mosk) 26:807-812(1992).
RN [9]
RP SEQUENCE OF 1009-1131 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93062884; PubMed=1435773;
RA Bliskovskii V.V., Berdichevskii F.B., Tkachenko A.V., Belova M.E.,
RA Chumakov I.M.;
RT "Coding part of the son gene small transcript contains four areas of
RT complete tandem repeats";

RL Mol. Biol. (Mosk) 26:793-806(1992).
RN [10]
RP SEQUENCE OF 1145-2426 FROM N.A. (ISOFORM F).
RX MEDLINE=93048367; PubMed=1424986;
RA Mattioni T., Hume C.R., Konigorski S., Hayes P., Osterweil Z.,
RA Lee J.S.;
RT "A cDNA clone for a novel nuclear protein with DNA binding
RT activity";
RL Chromosoma 101:618-624(1992).
RN [11]
RP SEQUENCE OF 1692-2175 FROM N.A. (ISOFORM A).
RX MEDLINE=89039788; PubMed=3054499;
RA Berdichevskii F.B., Chumakov I.M., Kiselev L.B.;
RT "Decoding of the primary structure of the son3 region in human
RT genome: identification of a new protein with unusual structure and
RT homology with DNA-binding proteins";
RL Mol. Biol. (Mosk) 22:794-801(1988).
RN [12]
RP SEQUENCE OF 1939-2426 FROM N.A. (ISOFORM J).
RC TISSUE=Cerebellum;
RX MEDLINE=99439804; PubMed=10509013;
RA Greenhalf W., Lee J., Chaudhuri B.;
RT "A selection system for human apoptosis inhibitors using yeast";
RL Yeast 15:1307-1321(1999).
CC -1- FUNCTION: Represses hepatitis B virus (HBV) core promoter activity
CC and transcription of HBV genes and production of HBV virions.
CC Binds to the consensus DNA sequence: 5'-GA[GT]AN[CG][AG]CC-3'.
CC Might protect cells from apoptosis. Might be involved in pre-mRNA
CC splicing (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear with a speckled distribution.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=10;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=F;
CC IsoId=P18583-1; Sequence=Displayed;
CC Name=A;
CC IsoId=P18583-2; Sequence=VSP_004401, VSP_004402, VSP_004403;
CC Name=B;
CC IsoId=P18583-3; Sequence=VSP_004404, VSP_004405;
CC Name=C;
CC IsoId=P18583-4; Sequence=VSP_004406, VSP_004407;
CC Name=D;
CC IsoId=P18583-5; Sequence=VSP_004408;
CC Name=E;
CC IsoId=P18583-6; Sequence=VSP_004409;
CC Name=G;
CC IsoId=P18583-7; Sequence=VSP_004410;
CC Name=H;
CC IsoId=P18583-8; Sequence=VSP_004411, VSP_004412;
CC Name=I;
CC IsoId=P18583-9; Sequence=VSP_004413;
CC Name=J;
CC IsoId=P18583-10; Sequence=VSP_004414, VSP_004415;
CC -1- TISSUE SPECIFICITY: Widely expressed, with the higher expression
CC seen in leukocyte and heart.
CC -1- DOMAIN: Contains 8 types of repeats which are distributed in 3
CC regions.
CC -1- MISCELLANEOUS: Colocalizes with the pre-mRNA splicing factor
CC SFR32/SC-35.
CC -1- SIMILARITY: Contains 1 G-patch domain.
CC -1- SIMILARITY: Contains 1 DBRM (double-stranded RNA-binding) domain.
CC -1- CAUTION: ISOFORM A SEQUENCE FROM REF.7 DIFFERS FROM THAT SHOWN DUE
CC TO A FRAMESHIFT.
CC -1- CAUTION: ISOFORM F SEQUENCE FROM REF.10 DIFFERS FROM THAT SHOWN
CC DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-ebc.ch/announce/>)

```

or send an email to license@isb-sib.ch).
-----
CC EMBL; AF380179; AAL34497.1; -.
CC EMBL; X63753; CAA45282.1; ALT_FRAME.
DR EMBL; M36428; AAA36624.1; -.
DR EMBL; AF380180; AAL34498.1; -.
DR EMBL; AF380181; AAL34499.1; -.
DR EMBL; AF380182; AAL34500.1; -.
DR EMBL; AF380183; AAL34501.1; -.
DR EMBL; AF380184; AAL34502.1; -.
DR EMBL; AY026895; AAK07692.1; -.
DR EMBL; AF435977; AAL30810.1; -.
DR EMBL; X63751; CAC59885.1; -.
DR EMBL; AB028942; BAA82971.1; -.
DR EMBL; X63071; CAA44793.1; ALT_FRAME.
DR EMBL; AF139897; AAD50078.1; -.
DR EMBL; AK024752; BAB14985.1; -.
DR EMBL; AF161428; AAF28988.1; -.
DR EMBL; AF161430; AAF28990.1; -.
DR Genew; HGNC:11183; SON.
DR GK; P18583; -.
DR MIM; 182465; -.
DR GO; GO:0008189; F.apoptosis inhibitor activity; IDA.
DR GO; GO:0003677; F.DNA binding activity; TAS.
DR GO; GO:0005916; P.anti-apoptosis; IDA.
DR InterPro; IPR001159; DS_RBD.
DR Pfam; PF00035; dxrm; 1.
DR Pfam; PF01585; G_patch; 1.
DR SMART; SM00443; G_patch; 1.
DR PROSITE; PS0137; DS_RBD; 1.
DR PROSITE; PS0174; G_PATCH; 1.
KW RNA-binding; DNA-binding; Nuclear protein; Repeat;
Alternative splicing.
FT DOMAIN 726 895 17 X 10 AA TANDEM REPEATS OF L-A-[ST]-
[NSG]-[TS]-WDSQM.
Query Match 68.8%; Score 75; DB 1; Length 2426;
Best Local Similarity 75.0%; Pred. No. 0.86;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKK 21
Db 109 KKKKKKKKKKKKKKKKKKK 128

RESULT 7
Y694_METJA STANDARD; PRT; 414 AA.
AC Q58105;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0694.
GN MJ0694.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus

```

```

jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U67516; AAB98689.1; -.
DR PIR; F64386; P64386.
DR TIGR; MJ0694; -.
DR InterPro; IPR002687; Nop.
DR Pfam; PF01798; Nop; 1.
DR ProDom; PD004104; Nop; 1.
KW Hypothetical protein; Complete proteome.
FT DOMAIN 349 414 ASP/GLU/LYS-RICH.
SQ SEQUENCE 414 AA; 47799 MW; A9092BFC3C82C407 CRC64;
Query Match 67.9%; Score 74; DB 1; Length 414;
Best Local Similarity 75.0%; Pred. No. 0.27;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKK 21
Db 377 KKKKKKKKKKKKKKKKKKK 396

RESULT 8
CBF5_KLUJA STANDARD; PRT; 474 AA.
AC O13473;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Centromere/microtubule binding protein CBF5 (Centromere-binding factor
DE 5) (Nucleolar protein CBF5).
GN CBF5.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JBD100;
RX MEDLINE=98144788; PubMed=9483794;
RA Winkler A.A., Bobok A., Zonneveld B.J.M., Steensma H.Y.,
RA Hooykaas P.J.J.;
RT "The lysine-rich C-terminal repeats of the centromere-binding factor
RT 5 (Cbfs) of Kluyveromyces lactis are not essential for function.";
RL Yeast 14:37-48(1998).
CC -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC SOME WAY ASSOCIATED WITH THE CBF3 110 kDa SUBUNIT (CBF3A) (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -!- SIMILARITY: Contains 1 PUA domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF008563; AAC64862.1; -.
DR InterPro; IPR004802; Cbf5.

```

DR InterPro; IPR002478; PUA.
 DR InterPro; IPR002501; TruB_N.
 DR InterPro; IPR004521; Unchar_dom_2.
 DR Pfam; PF01472; PUA; 1.
 DR Pfam; PF01509; TruB_N; 1.
 DR SMART; SMO0359; PUA; 1.
 DR TIGRFAMs; TIGR00425; CBF5; 1.
 DR TIGRFAMs; TIGR00451; unchar_dom_2; 1.
 DR PROSITE; PS00890; PUA; 1.
 KW Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
 FT DOMAIN 265 340 PUA.
 FT REPEAT 431 460 9 X 3 AA TANDEM REPEATS OF K-K-[DE].
 FT REPEAT 431 433 1.
 FT REPEAT 434 436 2.
 FT REPEAT 437 439 3.
 FT REPEAT 440 442 4.
 FT REPEAT 443 445 5.
 FT REPEAT 446 448 6.
 FT REPEAT 449 451 7.
 FT REPEAT 452 454 8.
 FT REPEAT 455 457 9.
 SQ SEQUENCE 474 AA; 53630 MW; 95306CE7FEA756C CRC64;
 Query Match 67.9%; Score 74; DB 1; Length 474;
 Best Local Similarity 70.0%; Pred. No. 0.3;
 Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
 QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 434 KKEKKEKKKKKKKKKKKK 453
 RESULT 9
 CBF5 YEAST
 ID -CBF5 YEAST STANDARD; PRT; 483 AA.
 AC P33322;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Centromere/microtubule binding protein CBF5 (Centromere-binding factor
 DE 5) (Nucleolar protein CBF5) [P64].
 GN CBF5 OR YLR175W OR L9470.11.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=93330283; PubMed=8336724;
 RA Jiang W., Middleton K., Yoon H.-J., Fouquet C., Carbon J.;
 RT "An essential yeast protein, CBF5p, binds in vitro to centromeres and
 RL microtubules." Mol. Cell. Biol. 13:4884-4893(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=97313267; PubMed=9169871;
 RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansoerge W.,
 RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
 RA Etian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
 RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
 RA Louis E.J., Messenguy F., Naves H.-W., Miesga T., Moestl D.,
 RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
 RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
 RA Scharfe M., Scherens B., Scholler P., Schwall C., Schwarz S.,
 RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
 RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
 RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII";
 RL Nature 387:90(1997).
 CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
 CC CENTROMERE DNA-CBF5-BINDING FACTOR AND IS INVOLVED IN MITOTIC
 CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN

CC SOME WAY ASSOCIATED WITH THE CBF3 110 kDa SUBUNIT (CBF3A).
 CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
 CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
 CC -1- SIMILARITY: Contains 1 PUA domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; LI2351; AAA34473.1; -.
 DR EMBL; U17246; AAB67463.1; -.
 DR PIR; S41853; S41853.
 DR SGD; S0004165; CBF5.
 DR GO; GO:0005732; C-small nucleolar ribonucleoprotein complex; RPI.
 DR InterPro; IPR004802; Cbf5.
 DR InterPro; IPR002478; PUA.
 DR InterPro; IPR002501; TruB_N.
 DR InterPro; IPR004521; Unchar_dom_2.
 DR Pfam; PF01472; PUA; 1.
 DR Pfam; PF01509; TruB_N; 1.
 DR SMART; SMO0359; PUA; 1.
 DR TIGRFAMs; TIGR00425; CBF5; 1.
 DR TIGRFAMs; TIGR00451; unchar_dom_2; 1.
 DR PROSITE; PS00890; PUA; 1.
 KW Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
 FT DOMAIN 266 341 PUA.
 FT REPEAT 434 463 10 X 3 AA TANDEM REPEATS OF K-K-[DE].
 FT REPEAT 434 436 1.
 FT REPEAT 437 439 2.
 FT REPEAT 440 442 3.
 FT REPEAT 443 445 4.
 FT REPEAT 446 448 5.
 FT REPEAT 449 451 6.
 FT REPEAT 452 454 7.
 FT REPEAT 455 457 8.
 FT REPEAT 458 460 9.
 FT REPEAT 461 463 10.
 SQ SEQUENCE 483 AA; 54704 MW; D356B39FDC32E2D CRC64;
 Query Match 67.9%; Score 74; DB 1; Length 483;
 Best Local Similarity 70.0%; Pred. No. 0.31;
 Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
 QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 434 KKEKKEKKKKKKKKKKKK 453
 RESULT 10
 BRD3 HUMAN
 ID -BRD3 HUMAN STANDARD; PRT; 726 AA.
 AC Q15059; Q92645;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bromodomain-containing protein 3 (RING3-like protein).
 GN BRD3 OR RING3L OR KIAA0043.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=96051398; PubMed=7584044;
 RA Nomura N., Nagase T., Miyajima N., Suzuki T., Tanaka A., Sato S.,
 RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
 RT "Prediction of the coding sequences of unidentified human genes. II.
 RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by


```

RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 1:223-229 (1994).
RN [2]
RP SEQUENCE OF 363-726 FROM N.A.
RX MEDLINE=98038990; PubMed=9373153;
RA Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowsdale J., Beck S.;
RT "Chromosomal localization, gene structure and transcription pattern of
the ORFX gene, a homologue of the MHC-linked RING3 gene."
RL Gene 200:177-183 (1997).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Contains 2 bromodomains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; D26362; BAA05393.1; -
DR InterPro; IPR001487; Bromodomain.
DR EMBL; Z81330; CAB03630.1; -
DR HSP; Q92831; I891.
DR Genew; HGNC:1104; BRD3.
DR MIM; 601541; -
DR GO; GO:0005634; C:nucleus; NAS.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS00144; BROMODOMAIN_2; 2.
KW Bromodomain; Repeat; Nuclear protein.
FT DOMAIN 56 115 BROMODOMAIN 1.
FT DOMAIN 326 398 BROMODOMAIN 2.
FT DOMAIN 487 555 LYS-RICH.
FT DOMAIN 676 725 SER-RICH.
FT CONFLICT 465 466 EL -> DV (IN REF. 2).
SQ SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;

Query Match 67.9%; Score 74; DB 1; Length 726;
Best Local Similarity 70.0%; Pred. No. 0.42;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKKKK 21
DB 489 KKKKKKKKKKKKKKKKKKK 508

RESULT 11
ID CBF5 CANAL STANDARD; PRT; 479 AA.
AC 043101;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Centromere/microtubule binding protein CBF5 (Centromere-binding factor
DE 5) (Nucleolar protein CBF5).
GN CBF5.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA Jiang W., Clifford J., Koltin Y.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND IS INVOLVED IN MITOTIC
CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC SOME WAY ASSOCIATED WITH THE CBF3 110 kDa SUBUNIT (CBF3A) (BY
CC SIMILARITY).

```

```

CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -!- SIMILARITY: Contains 1 PUA domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U59149; AAB94297.1; -
DR InterPro; IPR004802; Cbf5.
DR InterPro; IPR002478; PUA.
DR InterPro; IPR002501; TruB_N.
DR Pfam; PF01472; PUA; 1.
DR Pfam; PF01509; TruB_N; 1.
DR SMART; SM00359; PUA; 1.
DR TIGRFAMs; TIGR00425; CBF5; 1.
DR TIGRFAMs; TIGR00451; unchar_dom_2; 1.
DR PROSITE; PS0890; PUA; 1.
KW Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT DOMAIN 267 342 PUA.
SQ SEQUENCE 479 AA; 54321 MW; 3BAF5104E12C9EB6 CRC64;

Query Match 66.1%; Score 72; DB 1; Length 479;
Best Local Similarity 70.0%; Pred. No. 0.47;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKKKK 21
DB 432 KKKKKKKKKKKKKKKKKKK 451

RESULT 12
ID GARP PLAFF STANDARD; PRT; 678 AA.
AC P13816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutamic acid-rich protein precursor.
GN GARP.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89040048; PubMed=2903445;
RA Triglia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.;
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
RT acid-rich protein (GARP)".
RL Mol. Biochem. Parasitol. 31:199-202 (1988).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; J03998; AAA29605.1; -
DR PIR; A54514; A54514.
KW Repeat; Malaria; Antigen; Signal.
FT SIGNAL 1 25
FT DOMAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.
FT DOMAIN 120 164 15 X 3 AA TANDEM REPEATS OF K-K-X.
FT DOMAIN 372 416 9 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 417 441 5 X APPROXIMATE TANDEM REPEATS.

```

```
FT DOMAIN 576 604 POLY-GLU.
FT DOMAIN 605 653 7 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 654 663 POLY-GLU.
SQ SEQUENCE 678 AA; 80551 MW; 2A8F85606496EA9E CRC64;

Query Match 66.1%; Score 72; DB 1; Length 678;
Best Local Similarity 70.0%; Pred. No. 0.62;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
||:|||||:|||||:|||||
Db 132 KKEKKKKKKKKKKKKKKKKKK 151
||:|||||:|||||:|||||

RESULT 13
CG79 HUMAN STANDARD; PRT; 351 AA.
ID CG79_HUMAN PRT; 351 AA.
AC Q9Y388;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein CGI-79.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2072150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of novel human genes evolutionarily conserved in
RT Caenorhabditis elegans by comparative proteomics.";
RL Genome Res. 10:703-713(2000).
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF151837; AAD34074.1; -.
CC HSSP; P26368; 202F.
CC InterPro; IPR00504; RNA_rec_mot.
CC Pfam; PF00076; rrm; 1.
CC SMART; SM00360; RRM; 1.
CC PROSITE; PS0102; RRM; 1.
CC PROSITE; PS00030; RRM_RNP_1; 1.
KW Hypothetical protein; RNA-binding.
FT DOMAIN 36 114 RNA-BINDING (RRM).
SQ SEQUENCE 351 AA; 39675 MW; 7B6E882D6B192EBE CRC64;

Query Match 65.1%; Score 71; DB 1; Length 351;
Best Local Similarity 73.4%; Pred. No. 0.46;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 20
||:|||||:|||||:|||||
Db 156 KKHKKKKKKKKKKKKKKKKKK 174
||:|||||:|||||:|||||

RESULT 14
CNG1 HUMAN STANDARD; PRT; 686 AA.
ID CNG1_HUMAN PRT; 686 AA.
AC P29973; Q16279; Q16485;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNGI)
DE (Cyclic nucleotide-gated channel alpha 1) (Cyclic nucleotide-gated
DE channel, photoreceptor) (Cyclic-nucleotide-gated cation channel 1)
DE
```

```
DE (Rod photoreceptor cGMP-gated channel alpha subunit).
GN CNGA1 OR CNCG1 OR CNCG.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=92210603; PubMed=1372902;
RA Pittler S.J., Lee A.K., Altherr M.R., Howard T.A., Seldin M.F.,
RA Hurwitz R.L., Wasmuth J.J., Baehr W.;
RT "Primary structure and chromosomal localization of human and mouse
RT rod photoreceptor cGMP-gated cation channel.";
RL J. Biol. Chem. 267:6257-6262(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=92356211; PubMed=1379636;
RA Dhailan R.S., Macke J.P., Eddy R.L., Shows T.B., Reed R.R.,
RA Yau K.-W., Nathans J.;
RT "Human rod photoreceptor cGMP-gated channel: amino acid sequence,
RT gene structure, and functional expression.";
RL J. Neurosci. 12:3248-3256(1992).
RN [3]
RP SEQUENCE OF 313-573 FROM N.A.
RX MEDLINE=95175019; PubMed=7532814;
RA Distler M., Biel M., Flocke V., Hofmann F.;
RT "Expression of cyclic nucleotide-gated cation channels in non-sensory
RT tissues and cells.";
RL Neuropharmacology 33:1275-1282(1994).
RN [4]
RP VARIANT ARRP PHE-316, AND VARIANTS GLN-28 AND ASN-114.
RX MEDLINE=96036047; PubMed=7479749;
RA Dryja T.P., Finn J.T., Peng Y.-W., McGee T.L., Berson E.L., Yau K.-W.;
RT "Mutations in the gene encoding the alpha subunit of the rod
RT cGMP-gated channel in autosomal recessive retinitis pigmentosa.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:10177-10181(1995).
CC -1- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC COUPLED CASCADE USING CNGP AS SECOND MESSENGER. THIS PROTEIN CAN
CC BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO AN OPENING OF THE CATION
CC CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
CC PHOTORECEPTORS.
CC -1- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER. FORMS HETEROOLIGOMERIC
CC COMPLEX WITH CNG4.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.
CC -1- DISEASE: Defects in CNGA1 are a cause of autosomal recessive
CC retinitis pigmentosa (ARRP) [MIM:123825], a disease that leads to
CC degeneration of retinal photoreceptor cells.
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC FAMILY.
CC -1- DATABASE: NAME=Mutations of the CNGA1 gene;
CC NOTE=Retina International's Scientific Newsletter;
CC WWW="http://www.retina-international.com/sci-news/cngalmut.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M84741; AAA52010.1; ALT_INIT.
CC EMBL; S42457; AAB22778.1; -.
CC EMBL; S76062; AAD14206.1; -.
CC Genew; HGNC:2148; CNGA1.
CC MIM; 123825; -.
CC GO; GO:0005887; C:integral to plasma membrane; TAS.
CC GO; GO:0006832; P:small molecule transport; TAS.
CC GO; GO:0007601; P:vision; TAS.
CC InterPro; IPR000595; CNMP_binding.
```

DR InterPro; IPR005821; Ion trans.
 DR InterPro; IPR001622; K+channel_pore.
 DR Pfam; PF00027; cNMP_binding; 1.
 DR Pfam; PF00520; ion_trans; 1.
 DR SMART; SM00100; cNMP; 1.
 DR PROSITE; PS00888; cNMP_BINDING_1; 1.
 DR PROSITE; PS00889; cNMP_BINDING_2; 1.
 DR PROSITE; PS00442; cNMP_BINDING_3; 1.
 KW Ionic channel; Ion transport; cAMP-binding; Transmembrane;
 KW Multigene family; Vision; Disease mutation; Polymorphism;
 KW Retinitis pigmentosa.
 FT DOMAIN 1 160
 FT TRANSMEM 161 181
 FT DOMAIN 182 194
 FT TRANSMEM 195 213
 FT DOMAIN 214 237
 FT TRANSMEM 238 257
 FT DOMAIN 258 295
 FT TRANSMEM 296 318
 FT DOMAIN 319 370
 FT TRANSMEM 371 390
 FT DOMAIN 391 474
 FT TRANSMEM 475 495
 FT DOMAIN 496 686
 FT NP_BIND 493 605
 FT BINDING 542 542
 FT BINDING 557 557
 FT CARBOHYD 421 421
 FT VARIANT 28 28
 FT VARIANT 114 114
 FT VARIANT 316 316
 FT CONFLICT 46 46
 FT CONFLICT 85 85
 FT CONFLICT 146 147
 FT CONFLICT 539 539
 FT CONFLICT 677 678
 SQ SEQUENCE 686 AA; 79126 MW; E500D216FC97AF6 CRC64;
 Query Match 65.1%; Score 71; DB 1; Length 686;
 Best Local Similarity 70.0%; Pred. No. 0.78;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 124 KKKKKKKKKKKKKKKKKKK 143
 RESULT 15
 ID TCOP_HUMAN STANDARD; PRT; 1411 AA.
 AC Q13428; Q99408; Q99860;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Treacle protein (Treacher Collins syndrome protein).
 GN TCOP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96154183; PubMed=8563749;
 RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K.,
 RA Bonner C.A., Koprivnikar K., Wasmuth J.J.;
 FT "Positional cloning of a gene involved in the pathogenesis of
 FT Treacher Collins syndrome.";
 RL Nat. Genet. 12:130-136(1996).
 RN [2]
 RP SEQUENCE FROM N.A.

RX MEDLINE=97228900; PubMed=9074926;
 RA Dixon J., Edwards S.J., Anderson I., Brass A., Scambler P.J.,
 RA Dixon M.J.;
 RT "Identification of the complete coding sequence and genomic
 RT organization of the Treacher Collins syndrome gene.";
 RL Genome Res. 7:223-234(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97250498; PubMed=9096354;
 RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,
 RA Ashley J.A., Lovett M., Jabs E.W.;
 RT "TCOP1 gene encodes a putative nucleolar phosphoprotein that exhibits
 RT mutations in Treacher Collins syndrome throughout its coding
 RT region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).
 RN [4]
 RP VARIANTS LEU-439; VAL-810; VAL-1313 AND GLY-1355, AND VARIANT TCS
 RP ARG-53.
 RX MEDLINE=97195537; PubMed=9042910;
 RA Edwards S.J., Gladwin A.J., Dixon M.J.;
 RT "The mutational spectrum in Treacher Collins syndrome reveals a
 RT predominance of mutations that create a premature-termination
 RT codon.";
 RL Am. J. Hum. Genet. 60:515-524(1997).
 CC -!- DISEASE: Defects in TCOP1 are the cause of Treacher Collins
 CC syndrome (TCS) [MIM:154500]. TCS is an autosomal dominant disorder
 CC of craniofacial development that occurs with an incidence of
 CC 1/50,000 live births. The clinical features of TCS are bilaterally
 CC symmetrical and include: (1) abnormalities of the external ears,
 CC atresia of the external ear canals, and malformation of the middle
 CC ear ossicles, which may result in conductive hearing loss; (2)
 CC lateral downward sloping of palpebral fissures, frequently with
 CC colobomas of the lower eyelids; (3) hypoplasia of the mandible and
 CC zygomatic complex; (4) cleft palate.
 CC -!- SIMILARITY: Contains 1 Lish domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U40847; AAC50903.1; -;
 CC EMBL; U76366; AAC51181.1; -;
 CC EMBL; U84664; AAC51185.1; -;
 CC EMBL; U84640; AAC51185.1; JOINED.
 CC EMBL; U84641; AAC51185.1; JOINED.
 CC EMBL; U84642; AAC51185.1; JOINED.
 CC EMBL; U84643; AAC51185.1; JOINED.
 CC EMBL; U84644; AAC51185.1; JOINED.
 CC EMBL; U84645; AAC51185.1; JOINED.
 CC EMBL; U84646; AAC51185.1; JOINED.
 CC EMBL; U84647; AAC51185.1; JOINED.
 CC EMBL; U84648; AAC51185.1; JOINED.
 CC EMBL; U84649; AAC51185.1; JOINED.
 CC EMBL; U84650; AAC51185.1; JOINED.
 CC EMBL; U84651; AAC51185.1; JOINED.
 CC EMBL; U84652; AAC51185.1; JOINED.
 CC EMBL; U84653; AAC51185.1; JOINED.
 CC EMBL; U84654; AAC51185.1; JOINED.
 CC EMBL; U84655; AAC51185.1; JOINED.
 CC EMBL; U84656; AAC51185.1; JOINED.
 CC EMBL; U84657; AAC51185.1; JOINED.
 CC EMBL; U84658; AAC51185.1; JOINED.
 CC EMBL; U84659; AAC51185.1; JOINED.
 CC EMBL; U84660; AAC51185.1; JOINED.
 CC EMBL; U84661; AAC51185.1; JOINED.
 CC EMBL; U84662; AAC51185.1; JOINED.
 CC EMBL; U84663; AAC51185.1; JOINED.
 CC EMBL; U79659; AAB40722.1; -;
 CC EMBL; U79645; AAB40722.1; JOINED.

```

DR ENBL; U79646; AAB40722.1; JOINED.
DR ENBL; U79647; AAB40722.1; JOINED.
DR ENBL; U79648; AAB40722.1; JOINED.
DR ENBL; U79649; AAB40722.1; JOINED.
DR ENBL; U79650; AAB40722.1; JOINED.
DR ENBL; U79651; AAB40722.1; JOINED.
DR ENBL; U79652; AAB40722.1; JOINED.
DR ENBL; U79653; AAB40722.1; JOINED.
DR ENBL; U79654; AAB40722.1; JOINED.
DR ENBL; U79655; AAB40722.1; JOINED.
DR ENBL; U79656; AAB40722.1; JOINED.
DR ENBL; U79657; AAB40722.1; JOINED.
DR ENBL; U79658; AAB40722.1; JOINED.
DR Genew; HGNC:11654; TCOF1.
DR MIM; 606847; -.
DR MIM; 154500; -.
DR GO; GO:0005730; C:nucleolus; TAS.
DR GO; GO:0005215; P:transporter activity; TAS.
DR GO; GO:0001501; P:skeletal development; TAS.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR003993; treacle.
DR Pfam; PF03546; treacle; 3.
DR PRINTS; PR01503; TREACLE.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS00896; Lish; 1.
KW Disease mutation; Polymorphism.
FT DOMAIN 6 38
FT DOMAIN 89 97 LISH.
FT DOMAIN 204 207 POLY-GLU.
FT DOMAIN 616 619 POLY-SER.
FT DOMAIN 919 924 POLY-SER.
FT DOMAIN 1285 1289 POLY-SER.
FT DOMAIN 1375 1386 POLY-LYS.
FT DOMAIN 1398 1405 POLY-LYS.
FT VARIANT 53 53 W -> R (in TCS).
FT VARIANT 439 439 /FTid=VAR_005630.
FT VARIANT 810 810 P -> L.
FT VARIANT 1313 1313 /FTid=VAR_005631.
FT VARIANT 1355 1355 A -> V.
FT VARIANT 1312 1312 /FTid=VAR_005632.
FT VARIANT 1355 1355 A -> V (in dbSNP:15251).
FT VARIANT 1312 1312 /FTid=VAR_005633.
FT VARIANT 1355 1355 D -> G.
FT CONFLICT 1312 1312 /FTid=VAR_005634.
FT CONFLICT 1411 AA; 144312 MW; 3880203D985C2699 CRC64; K -> Q (IN REF. 2).
SQ SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64;

Query Match 64.7%; Score 70.5; DB 1; Length 1411;
Best Local Similarity 54.8%; Pred. No. 1.5;
Matches 17; Conservative 2; Mismatches 1; Indels 11; Gaps 1;

QY 2 KKKKKKKKKKKKKKKKKKKKKKKKKKKKK 21
Db 1375 KKKKKKKKKKKKKKKKKKKKKKKKKKKKK 1405

Search completed: January 30, 2004, 00:20:44
Job time : 5.73239 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:52 ; Search time 21.6901 Seconds
(without alignments)
249.842 Million cell updates/sec

Title: US-09-461-684C-1
Perfect score: 109
Sequence: 1 CKKKKKKKKKKKKKKKKKKK 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	91.7	55	4 Q8N6F0	Q8n6f0 homo sapien
2	100	91.7	80	10 Q8S7D3	Q8s7d3 oryza sativ
3	100	91.7	113	10 Q8LQP6	Q8lqp6 oryza sativ
4	100	91.7	128	3 Q9P529	Q9p529 neurospora
5	100	91.7	129	11 Q35807	Q35807 rattus norv
6	100	91.7	168	4 Q9H5V6	Q9h5v6 homo sapien
7	100	91.7	206	5 Q8I247	Q8i247 plasmodium
8	100	91.7	215	11 Q64075	Q64075 rattus sp.
9	100	91.7	260	10 Q9L6Z9	Q9l6z9 arabidopsis
10	100	91.7	380	4 Q9NT34	Q9nt34 homo sapien
11	100	91.7	515	5 Q8SWR7	Q8swr7 drosophila
12	100	91.7	517	10 Q9LXR2	Q9lxr2 arabidopsis
13	100	91.7	531	6 Q951V6	Q951v6 macaca fasc
14	100	91.7	720	4 Q9H6Q7	Q9h6q7 homo sapien
15	100	91.7	791	5 Q8T2U7	Q8t2u7 dictyosteli
16	97	89.0	667	4 Q9HC48	Q9hc48 homo sapien

17	96	88.1	476	10 Q8RVV4	Q8rvv4 lycopersico
18	95	87.2	489	5 Q8T3H4	Q8t3h4 drosophila
19	95	87.2	658	11 Q8CGI8	Q8cgi8 mus musculus
20	92	84.4	4433	5 Q8IJ15	Q8ij15 plasmodium
21	91	83.5	3351	5 Q8IBK4	Q8ibk4 plasmodium
22	90	82.6	39	5 Q8MSJ8	Q8msj8 drosophila
23	90	82.6	257	4 Q9H5Y3	Q9h5y3 homo sapien
24	90	82.6	686	4 Q9NMF0	Q9nmf0 homo sapien
25	90	82.6	2820	5 Q8IM32	Q8im32 plasmodium
26	89	81.7	83	11 Q9ER82	Q9er82 mus musculus
27	89	81.7	1710	5 Q8I239	Q8i239 plasmodium
28	88	80.7	109	5 Q8I3G8	Q8i3g8 plasmodium
29	87	79.8	107	10 Q9LQF6	Q9lqf6 arabidopsis
30	87	79.8	476	5 Q8IFN3	Q8ifn3 plasmodium
31	86	78.9	102	11 Q8C5Y8	Q8c5y8 mus musculus
32	86	78.9	741	5 Q8IPR6	Q8ipr6 drosophila
33	85	78.0	978	5 Q8IB83	Q8ib83 plasmodium
34	84	77.1	339	4 Q8IVS7	Q8ivs7 homo sapien
35	83	76.1	227	4 Q8TA86	Q8ta86 homo sapien
36	83	76.1	1337	5 Q8IC23	Q8ic23 plasmodium
37	82.5	75.7	710	5 Q8IDM7	Q8idm7 plasmodium
38	82	75.2	943	5 Q8ILY0	Q8ily0 plasmodium
39	81	74.3	253	11 Q8CFH9	Q8cfh9 mus musculus
40	81	74.3	547	4 Q8IV81	Q8iv81 homo sapien
41	81	74.3	1082	5 Q8ISY6	Q8isy6 plasmodium
42	81	74.3	1313	5 Q8IKC0	Q8ikc0 plasmodium
43	81	74.3	3193	5 Q8IS90	Q8is90 plasmodium
44	81	74.3	4550	5 Q77336	Q77336 plasmodium
45	80	73.4	213	11 P97762	P97762 mus musculus

ALIGNMENTS

RESULT 1
Q8N6F0 PRELIMINARY; PRT; 55 AA.
AC Q8N6F0; 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to LOC201361.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC030525; AAH30525.1;
SQ SEQUENCE 55 AA; 7251 MW; 0906032B284006BA CRC64;

Query Match 91.7%; Score 100; DB 4; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.9e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
Db 21 KKKKKKKKKKKKKKKKKKK 40

RESULT 2
Q8S7D3 PRELIMINARY; PRT; 80 AA.
AC Q8S7D3; 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Hypothetical 9.4 kDa protein.
GN OSJNEA0057L21.23.
OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
RN [1] NCBI_TaxID=4530;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA Tsitrin T., Riggs F., Hsiao J., Ziemann V., Blunt S., Pai G.,
RA VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA Salzberg S.L., White O., Fraser C.M.,
RT "Oryza sativa chromosome 10 BAC OSUNBa0057L21 genomic sequence."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC087599; AAL79706.1; -
DR Gramene; Q857D3; -
KW Hypothetical protein.
SQ SEQUENCE 80 AA; 9362 MW; 0177C863133B21D8 CRC64;

Query Match 91.7%; Score 100; DB 10; Length 80;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKK 21
DB 48 KKKKKKKKKKKKKKKKK 67

RESULT 3

ID Q8LQ6 PRELIMINARY; PRT; 113 AA.
AC Q8LQ6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE OJ1117_G01.13 protein.
GN OJ1117_G01.13
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
RN [1] NCBI_TaxID=39947;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone:OJ1117_G01."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003374; BAB93330.1; -
DR Gramene; Q8LQ6; -
SQ SEQUENCE 113 AA; 13660 MW; 597DB0E2B2AA3EF CRC64;

Query Match 91.7%; Score 100; DB 10; Length 113;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKK 21
DB 10 KKKKKKKKKKKKKKKKK 29

RESULT 4

ID Q9P529 PRELIMINARY; PRT; 128 AA.
AC Q9P529;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 15.2 kDa protein.
GN B24H17.160.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1] SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RA Schulte U., Algen V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356845; CAB92638.2; -
KW Hypothetical protein.
SQ SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFB70765 CRC64;

Query Match 91.7%; Score 100; DB 3; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKK 21
DB 71 KKKKKKKKKKKKKKKKK 90

RESULT 5

ID O35807 PRELIMINARY; PRT; 129 AA.
AC O35807;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE MICROVASCULAR endothelial differentiation protein 2.
GN MDG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Epithelium;
RX MEDLINE=98172708; PubMed=9511718;
RA Proels F., Loser B., Marx M.;
RT "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA
species, during in vitro angiogenesis."
RL Exp. Cell Res. 239:1-10(1998).
DR EMBL; Y08769; CAA70022.1; -
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00069; Pkinase; 1.
DR PRODOM; PD000001; Prot_kinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 129 AA; 15080 MW; 38102272BEE2EDB4 CRC64;

Query Match 91.7%; Score 100; DB 11; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKK 21
DB 85 KKKKKKKKKKKKKKKKK 104

RESULT 6

ID Q9H5V6 PRELIMINARY; PRT; 168 AA.
AC Q9H5V6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ22976 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Iseigai T., Sugano S.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026629; BAB15513.1; -.
KW Hypothetical protein.
FT NON_TER 168
SQ SEQUENCE 168 AA; 19549 MW; A19DED195F8A1A90 CRC64;

Query Match 91.7%; Score 100; DB 4; Length 168;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 140 KKKKKKKKKKKKKKKKKKK 159

RESULT 7
Q81247
ID Q81247 PRELIMINARY; PRT; 206 AA.
AC Q81247
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
DE PFA0475C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berrian M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kethornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL031745; CAD49055.1; -.
KW Hypothetical protein.
SQ SEQUENCE 206 AA; 25047 MW; 1192E49A3DC4523F CRC64;

Query Match 91.7%; Score 100; DB 5; Length 206;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 185 KKKKKKKKKKKKKKKKKKK 204

RESULT 8
Q64075
ID Q64075 PRELIMINARY; PRT; 215 AA.
AC Q64075;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
```

```
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nucleoporin p62 homolog protein (Fragment).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95151924; PubMed=7849178;
RA Wang Z.Q., Akmal K.M., Kim K.H.;
RT "An unusual nucleoporin-related messenger ribonucleic acid is present
RT in the germ cells of rat testis.";
RL Biol. Reprod. 51:1022-1030(1994).
DR EMBL; S75997; AAB33384.1; -.
KW Porin.
FT NON_TER 1
SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD88 CRC64;

Query Match 91.7%; Score 100; DB 11; Length 215;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 35 KKKKKKKKKKKKKKKKKKK 54

RESULT 9
Q9LGZ9
ID Q9LGZ9 PRELIMINARY; PRT; 260 AA.
AC Q9LGZ9
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: F1D9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Nakamura Y.;
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002460; BAA97098.1; -.
DR InterPro; IPR005819; Histone_H5.
DR PRINTS; PR00624; HISTONEH5.
SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;

Query Match 91.7%; Score 100; DB 10; Length 260;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KKKKKKKKKKKKKKKKKKKKK 21
Db 7 KKKKKKKKKKKKKKKKKKK 26

RESULT 10
Q9NT34
ID Q9NT34 PRELIMINARY; PRT; 380 AA.
AC Q9NT34
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP434I1120.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```

RN  SEQUENCE FROM N.A.
RP  Tissue=Testis;
RC  Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RL  EMBL; AL137556; CAB70810.1; -.
DR  Genew; HGNC:15736; CL7orf28.
KW  Hypothetical protein.
FT  NON TER 380
SQ  SEQUENCE 380 AA; 42689 MW; 67F50DD101346AFB CRC64;

Query Match          91.7%; Score 100; DB 4; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.00028;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2 KKKKKKKKKKKKKKKKKKKKKKK 21
DB  355 KKKKKKKKKKKKKKKKKKKKKKK 374

RESULT 11
Q8SWR7
ID  Q8SWR7 PRELIMINARY; PRT; 515 AA.
AC  Q8SWR7;
DT  01-JUN-2002 (TrEMBLrel. 21, Created)
DT  01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE  GH22607p (Fragment).
GN  CG7180.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Berkely;
RA  Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA  Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA  George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA  Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA  Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA  Celniker S.;
RL  Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AY095518; AAM12251.1; -.
DR  FlyBase; FBgn0032673; CG7180.
DR  InterPro; IPR000387; TYR phosphatase.
DR  InterPro; IPR000242; TYR PP.
DR  Pfam; PF00102; Y phosphatase; 1.
DR  SMART; SM00194; PTPC; 1.
DR  PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR  PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR  PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW  Hydrolyase.
FT  NON TER 515
SQ  SEQUENCE 515 AA; 59080 MW; B2825B7EEA96195E CRC64;

Query Match          91.7%; Score 100; DB 5; Length 515;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2 KKKKKKKKKKKKKKKKKKKKKKK 21
DB  493 KKKKKKKKKKKKKKKKKKKKKKK 512

RESULT 12
Q9LXR2
ID  Q9LXR2 PRELIMINARY; PRT; 517 AA.
AC  Q9LXR2;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

```

```

DE  Hypothetical 59.7 kDa protein.
GN  T2ON10.250.
OS  Arabidopsis thaliana (Mouse-ear cress).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC  eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX  NCBI_TaxID=3702;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA  Rudd S., Lemcke K., Mayer K.F.X., Quetier P., Salanoubat M.;
RL  Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RA  EU Arabidopsis sequencing project;
RL  Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AL353032; CAB88307.1; -.
DR  InterPro; IPR001810; F-box.
DR  InterPro; IPR006566; FBD.
DR  Pfam; PF00646; F-box; 1.
DR  SMART; SM00579; FBD; 1.
DR  SMART; SM00256; FBOX; 1.
DR  PROSITE; PS50181; FBOX; 1.
KW  Hypothetical protein.
SQ  SEQUENCE 517 AA; 59689 MW; EC6D957D01F86E70 CRC64;

Query Match          91.7%; Score 100; DB 10; Length 517;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2 KKKKKKKKKKKKKKKKKKKKKKK 21
DB  444 KKKKKKKKKKKKKKKKKKKKKKK 463

RESULT 13
Q95LV6
ID  Q95LV6 PRELIMINARY; PRT; 531 AA.
AC  Q95LV6;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT  01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE  Hypothetical 61.4 kDa protein (Fragment).
OS  Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC  Cercopithecinae; Macaca.
OX  NCBI_TaxID=9541;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  Tissue=Testis;
RA  Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA  Terao K., Sugano S.;
RT  "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT  libraries."
RL  Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AB071085; BAB64479.1; -.
KW  Hypothetical protein.
FT  NON TER 531
SQ  SEQUENCE 531 AA; 61389 MW; B55996B4F5CDD60C CRC64;

Query Match          91.7%; Score 100; DB 6; Length 531;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2 KKKKKKKKKKKKKKKKKKKKKKK 21
DB  502 KKKKKKKKKKKKKKKKKKKKKKK 521

RESULT 14
Q9H6Q7
ID  Q9H6Q7 PRELIMINARY; PRT; 720 AA.

```


AC Q9H607;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ21979 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AK025632; BAB15196.1; -
KW Hypothetical protein.
FT NON_TER 720 720
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;
Query Match 91.7%; Score 100; DB 4; Length 720;
Best Local Similarity 100.0%; Pred. No. 0.00043;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 KKKKKKKKKKKKKKKKK 21
DB 692 KKKKKKKKKKKKKKKKK 711
RESULT 15
OBT2U7 PRELIMINARY; PRT; 791 AA.
AC OBT2U7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 92.4 kDa protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.P., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AC115574; AAL92183.1; -
DR InterPro; IPR005033; YEATS.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF03366; YEATS; 1.
DR SMART; SM00355; Znf_C2H2; 1.
KW Hypothetical protein.
SQ SEQUENCE 791 AA; 92375 MW; D66CCBDEC92352C CRC64;
Query Match 91.7%; Score 100; DB 5; Length 791;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 KKKKKKKKKKKKKKKKK 21
DB 769 KKKKKKKKKKKKKKKKK 788

Search completed: January 30, 2004, 00:24:37
Job time : 22.6901 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 00:13:12 ; Search time 10.6808 Seconds
 (without alignments)
 225.098 Million cell updates/sec

Title: US-09-461-684C-2

Perfect score: 109

Sequence: 1 CFAAAAAAEEEEAAAAA 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	77	70.6	265	2 S19113	cgcr-4 protein - C
2	72	66.1	165	2 B87702	ribosomal protein
3	71	65.1	205	2 S19114	cgcr-1 protein - C
4	70	64.2	183	2 S24960	gene C98 protein -
5	70	64.2	401	2 A48423	engrailed homeodom
6	69	63.3	97	2 S02376	antifreeze protein
7	69	63.3	1028	2 A56038	DNA-binding protein
8	69	63.3	1213	2 S16356	ovo protein - fru
9	69	63.3	1668	2 T13748	sex comb protein -
10	68	62.4	403	2 A81882	probable dihydrol
11	68	62.4	655	2 A28945	neurogenesis regul
12	68	62.4	873	2 B52225	ecdysone-induced p
13	68	62.4	2023	2 T13154	polycomb protein e
14	67	61.5	314	2 JCS273	hypothetical prote
15	67	61.5	460	2 T33110	hypothetical prote
16	67	61.5	494	2 A42170	zinc finger protei
17	67	61.5	497	2 JCS076	myc-associated zin
18	67	61.5	604	2 A39369	homeotic protein B
19	67	61.5	606	2 S13367	Om (ID) protein - f
20	67	61.5	627	2 T02610	probable YME1 ATP-
21	67	61.5	671	2 C96534	probable Poly-A Bi
22	67	61.5	1226	2 T24045	hypothetical prote
23	67	61.5	2715	2 T13049	eyelid - fruit fly
24	66.5	61.0	543	2 B39369	homeotic protein B
25	65.5	60.1	273	2 T51010	hypothetical prote
26	65	59.6	85	1 FDL4W	antifreeze protei
27	65	59.6	91	2 A22592	antifreeze protei
28	65	59.6	333	2 A39065	homeotic protein B
29	65	59.6	475	2 A43915	homeotic protein e

30	65	59.6	644	2 S39356	transcription fact
31	65	59.6	703	2 T48600	kinase-like protei
32	64	58.7	109	1 R6UTP1	acidic ribosomal p
33	64	58.7	392	2 B48423	homeotic protein e
34	64	58.7	1065	2 T13230	dachshund isoform
35	64	58.7	1072	2 T13232	dachshund protein
36	64	58.7	1074	2 T13229	dachshund protein
37	64	58.7	1081	2 T13231	dachshund protein
38	64	58.7	1533	2 A46221	abdominal segment
39	64	58.7	2038	2 A43742	female sterile hom
40	63	57.8	179	2 AF2908	50S ribosomal prot
41	63	57.8	179	2 F97683	50S ribosomal prot
42	63	57.8	513	2 A48233	polyomavirus enhan
43	63	57.8	568	2 T39675	asparaginyl-trna s
44	63	57.8	581	2 E75383	conserved hypothet
45	63	57.8	846	2 S52418	GTP-binding regula

ALIGNMENTS

RESULT 1

S19113
 cgcr-4 protein - Chlamydomonas reinhardtii (fragment)
 C:Species: Chlamydomonas reinhardtii
 C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S19113; S14466
 R:Wakarchuk, W.W.; Mueller, F.W.; Beck, C.F.
 Plant Mol. Biol. 18, 143-146, 1992
 A:Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements
 A:Reference number: S19113; NUID:92119224; PMID:1731966
 A:Accession: S19113
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-265 <WAK>
 A:Cross-references: ENBL:X17208; NID:g18136; PIDN:CAA35080.1; PID:g18137
 C:Genetics:
 A:Gene: cgcr-4

Query Match 70.6%; Score 77; DB 2; Length 265;
 Best Local Similarity 82.6%; Pred. No. 0.25;
 Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAEEEEAAAAA 25
 DB 154 AAAAAAEEAAKARVAAEAAAA 176
 |||||:|||||

RESULT 2

B87702
 ribosomal protein S16 [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: B87702
 R:Nierman, W.C.; Feldlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; NUID:21173698; PMID:11259647
 A:Accession: B87702
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-165 <STO>
 A:Cross-references: GB:AE005673; NID:g13425408; PIDN:AAK25614.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC3652

Query Match 66.1%; Score 72; DB 2; Length 165;
 Best Local Similarity 75.0%; Pred. No. 0.56;
 Matches 18; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAAAAAEEEEAAAAA 25

C;Accession: A56038
R;Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
Mol. Cell. Biol. 14, 6809-6818, 1994
A;Title: Multiple products from the shavenbaby-ovo gene region of *Drosophila melanogaster*
A;Reference number: A56038; MUID:95021209; PMID:7935398
A;Accession: A56038
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1028 <GAR>
A;Cross-references: GB:U11383; NID:G520526; PIDN:AAB60216.1; PID:G520527
C;Genetics:
A;Gene: ovo
A;Cross-references: FlyBase:FBgn0003028

Query Match 63.3%; Score 69; DB 2; Length 1028;
Best Local Similarity 85.7%; Pred. No. 4.2;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAEEEEAAAAAEEAAA 23
||||| ||||| ||||| ||||| |||||
Db 497 AAAAAAEEEEAAAAAEEAAA 517

RESULT 8
S16356
Ovo protein - fruit fly (*Drosophila melanogaster*)
C;Species: *Drosophila melanogaster*
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
C;Accession: S16356
R;Nevel-Ninio, M.; Terracoli, R.; Kafatos, F.C.
EMBO J. 10, 2259-2266, 1991
A;Title: The ovo gene of *Drosophila* encodes a zinc finger protein required for female germline development
A;Reference number: S16356; MUID:91293102; PMID:1712294
A;Accession: S16356
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1213 <MEV>
A;Cross-references: EMBL:X59772
C;Genetics:
A;Gene: FlyBase:ovo
A;Cross-references: FlyBase:FBgn0003028
A;Introns: 931/3; 1152/3

Query Match 63.3%; Score 69; DB 2; Length 1213;
Best Local Similarity 85.7%; Pred. No. 4.7;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAEEEEAAAAAEEAAA 23
||||| ||||| ||||| ||||| |||||
Db 860 AAAAAAEEEEAAAAAEEAAA 880

RESULT 9
T13748
Sex comb protein - fruit fly (*Drosophila melanogaster*)
C;Species: *Drosophila melanogaster*
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13748
R;Sinclair, D.A.R.; Milne, T.A.; Hodgson, J.W.; Shellard, J.; Salinas, C.A.; Kyba, M.; Ravel, J.; et al.
Development 125, 1207-1216, 1998
A;Title: The additional sex combs gene of *Drosophila* encodes a chromatin protein that binds to the sex comb enhancer
A;Reference number: Z17750; MUID:98146384; PMID:9477319
A;Accession: T13748
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1668 <SIN>
A;Cross-references: EMBL:AJ001164; NID:G3292938; PIDN:CAA04568.1; PID:G3292939
C;Genetics:
A;Cross-references: FlyBase:FBgn0000142
A;Function:
A;Description: involved in repression of homeotic loci

Query Match 63.3%; Score 69; DB 2; Length 1668;

Best Local Similarity 78.3%; Pred. No. 6;
Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAAAAAEEEEAAAAAEEAAA 24
||||| ||||| ||||| ||||| |||||
Db 128 KAAAAAEEEEAAAAAEEAAA 150

RESULT 10
A81882
Probable dihydrolipoamide S-succinyltransferase (EC 2.3.1.61) E2 component NMA1150 [impor
C;Species: *Neisseria meningitidis*
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: A81882
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: A81882
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-403 <PAR>
A;Cross-references: GB:AL162755; GB:AL157959; NID:G7379742; PIDN:CA8844412.1; PID:G737984;
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: sucB; NMA1150
C;Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology
C;Keywords: acyltransferase; coenzyme A

Query Match 62.4%; Score 68; DB 2; Length 403;
Best Local Similarity 75.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 EAAAAAEEEEAAAAAEEAAA 25
||||| ||||| ||||| ||||| |||||
Db 85 EAPAAATAAAEAPAAAPAAAPAA 108

RESULT 11
A29945
neurogenesis regulatory protein - fruit fly (*Drosophila melanogaster*) (fragment)
N;Alternate names: single-minded gene protein
C;Species: *Drosophila melanogaster*
C;Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Mar-1998
C;Accession: A29945
R;Crews, S.T.; Thomas, J.B.; Goodman, C.S.
Cell 52, 143-151, 1988
A;Title: The *Drosophila* single-minded gene encodes a nuclear protein with sequence simil
A;Reference number: A29945; MUID:88151023; PMID:3345560
A;Accession: A29945
A;Molecule type: mRNA
A;Residues: 1-655 <CRE>
A;Cross-references: GB:M19020; NID:gl58464; PID:gl58465
C;Genetics:
A;Gene: sim
A;Cross-references: FlyBase:FBgn0004666
C;Keywords: DNA binding; transcription regulation

Query Match 62.4%; Score 68; DB 2; Length 655;
Best Local Similarity 66.7%; Pred. No. 3.8;
Matches 16; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAAAAAEEEEAAAAAEEAAA 25
||||| ||||| ||||| ||||| |||||
Db 366 QAAQAQAQAQAQAQAQAQAQA 389

RESULT 12
B53225
ecdysone-induced protein E74A - fruit fly (*Drosophila virilis*)
C;Species: *Drosophila virilis*
C;Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 21-Feb-1997

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:27 ; Search time 5.6338 Seconds
(without alignments)
208.681 Million cell updates/sec

Title: US-09-461-684C-2

Perfect score: 109

Sequence: 1 CEAAAAAEEEEAAAAEEEEAAAAA 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	72.5	332	1 FA9A_HUMAN	Q8121 homo sapien
2	75.5	69.3	518	1 TPM4_DROME	P49455 drosophila
3	72	66.1	165	1 RS16_CAUCR	P58122 caulobacter
4	70	64.2	183	1 OLEC_BRANA	P29526 brassica na
5	70	64.2	401	1 HME1_MOUSE	P09065 mus musculus
6	69	63.3	97	1 ANP_LIMFE	P09031 limanda fer
7	69	63.3	1028	1 OVO_DROME	P51521 drosophila
8	69	63.3	1669	1 ASX_DROME	Q9V727 drosophila
9	68	62.4	697	1 SIM_DROME	P05709 drosophila
10	68	62.4	1073	1 HR38_DROME	P49869 drosophila
11	67	61.5	314	1 PMXB_HUMAN	Q99453 homo sapien
12	67	61.5	314	1 PMXB_MOUSE	Q35690 mus musculus
13	67	61.5	477	1 MAZ_HUMAN	P56270 homo sapien
14	67	61.5	606	1 HM1D_DROAN	P22544 drosophila
15	66	60.6	386	1 HXAD_MOUSE	Q62424 mus musculus
16	66	60.6	388	1 HXAD_HUMAN	P31271 homo sapien
17	65	59.6	85	1 ANP4_PSEAM	P02734 pseudopleur
18	65	59.6	91	1 ANPX_PSEAM	F07835 pseudopleur
19	65	59.6	276	1 SX21_HUMAN	Q9V651 homo sapien
20	65	59.6	280	1 SX21_CHICK	Q9W755 gallus gall
21	65	59.6	475	1 EVX2_MOUSE	P49749 mus musculus
22	65	59.6	476	1 EVX2_HUMAN	Q03828 homo sapien
23	65	59.6	644	1 BTD_DROME	Q24266 drosophila
24	64	58.7	109	1 RLAI_TRYCR	P26643 trypanosoma
25	64	58.7	392	1 HME1_HUMAN	Q05925 homo sapien
26	64	58.7	1533	1 PUM_DROME	P25822 drosophila
27	64	58.7	2038	1 FSH_DROME	P13709 drosophila
28	63	57.8	179	1 RL19_AGT5	Q8ubz5 agrobacteri
29	63	57.8	376	1 FXE1_HUMAN	Q00358 homo sapien
30	63	57.8	521	1 RUN2_HUMAN	Q13950 h runt-rela
31	63	57.8	562	1 ARX_HUMAN	Q96g83 homo sapien
32	63	57.8	590	1 HMAA_DROME	P29555 drosophila
33	63	57.8	607	1 RUN2_MOUSE	Q08775 m runt-rela

34	63	57.8	1095	1 PIPA_DROME	P13217 drosophila
35	62	56.9	364	1 NK61_MESAU	Q60554 mesocricetu
36	62	56.9	365	1 NK61_MOUSE	Q99ma9 mus musculus
37	62	56.9	365	1 NK61_RAT	Q35762 rattus norv
38	62	56.9	1355	1 SALM_DROME	P39770 drosophila
39	61	56.0	31	1 ANP3_PAGBO	P02732 pagothemia
40	61	56.0	91	1 ANPY_PSEAM	P23699 pseudopleur
41	61	56.0	153	1 RS16_BIPLO	Q89791 bifidobacte
42	61	56.0	308	1 ABFI_DROME	P39413 drosophila
43	61	56.0	376	1 FXL2_HUMAN	P58012 homo sapien
44	61	56.0	421	1 PO41_MOUSE	P17208 mus musculus
45	61	56.0	423	1 PO41_HUMAN	Q01851 homo sapien

ALIGNMENTS

RESULT 1
FA9A_HUMAN STANDARD; PRT; 332 AA.
AC Q8121;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein FAM9A.
GN FAM9A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
RX MEDLINE=22202142; PubMed=12213195;
RA Martinez-Garay I., Jablonka S., Sutajova M., Steuernagel P., Gal A.,
RA Kutsche K.;
RT "A new gene family (FAM9) of low-copy repeats in Xp22.3 expressed
RT exclusively in testis: Implications for recombinations in this
RT region.";
RL Genomics 80:259-267(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- TISSUE SPECIFICITY: Expressed exclusively in testis.
CC -!- SIMILARITY: Belongs to the FAM9 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF494343; AAN07162.1; -
CC Genew; HGNC:18403; FAM9A.
DR Genew; HGNC:18403; FAM9A.
KW Nuclear protein.
FT DOMAIN 180 275 GLU-RICH.
FT DOMAIN 194 214 POLY-ALA.
FT DOMAIN 252 258 POLY-GLY.
SQ SEQUENCE 332 AA; 37339 MW; 92F22EC36038229C CRC64;
Query Match 72.5%; Score 79; DB 1; Length 332;
Best Local Similarity 83.3%; Pred. No. 0.066; 4; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 EAAAAAEEEEAAAAEEEEAAAAA 25
Db 190 EAEAEAEAEAEAEAEAEAEAEAEAE 213
RESULT 2
TPM4_DROME STANDARD; PRT; 518 AA.
ID TPM4_DROME
AC P49455; P49456; Q24425; Q24426;
DT 01-FEB-1996 (Rel. 33, Created)

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AE006023; AA25614.1; -
DR PIR; B87702; B87702.
DR HSP; P80379; IEMW.
DR TIGR; CC3652; -
DR HAMAP; MF 00385; -; 1.
DR InterPro; IPR000307; Ribosomal_S16.
DR Pfam; PF00886; Ribosomal_S16; 1.
DR ProDom; PD003791; Ribosomal_S16; 1.
DR TIGRFAMs; TIGR00002; S16; 1.
DR PROSITE; PS00732; RIBOSOMAL_S16; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 165 AA; 17605 MW; ED46FC2798C5BE1C CRC64;

Query Match 66.1%; Score 72; DB 1; Length 165;
Best Local Similarity 75.0%; Pred. No. 0.19;
Matches 18; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAAAAAEEAAAAAEEAAAAA 25
DB 115 QAEADAKAAAEKAAAAA 138

RESULT 4

ID_OLEC_BRANA STANDARD; PRT; 183 AA.
AC P29526;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Oleosin C98 (Fragment).
GN C98.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Anther;
RX MEDLINE=93386188; PubMed=8374615;
RA Roberts M.R., Hodge R., Ross J.H.E., Sorensen A., Murphy D.J.,
RA Draper J., Scott R.;
RT "Characterization of a new class of oleosins suggests a male
gametophyte-specific lipid storage pathway.";
RL Plant J. 3:629-636(1993).
CC -1- FUNCTION: MAY HAVE A STRUCTURAL ROLE TO STABILIZE THE LIPID BODY
DURING DESSICATION OF THE SEED BY PREVENTING COALESCENCE OF THE
OIL. PROBABLY INTERACTS WITH BOTH LIPID AND PHOSPHOLIPID MOIETIES
OF LIPID BODIES. MAY ALSO PROVIDE RECOGNITION SIGNALS FOR SPECIFIC
LIPASE ANCHORAGE IN LIPOLYSIS DURING SEEDLING GROWTH.
CC -1- SUBCELLULAR LOCATION: SURFACE OF OIL BODIES. OLEOSINS EXIST AT A
MONOLAYER LIPID/WATER INTERFACE.
CC -1- TISSUE SPECIFICITY: SPECIFIC TO THE MALE GAMETOPHYTE.
CC -1- SIMILARITY: BELONGS TO THE OLEOSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67142; CAA47623.1; -
DR PIR; S24960; S24960.
DR InterPro; IPR000136; Oleosin.

DR Pfam; PF01277; Oleosin; 1.
DR PROSITE; PS00811; OLEOSINS; 1.
KW Seed; Oil body; Multigene family.
FT NON_TER 1 1
FT DOMAIN <1 23 POLAR.
FT DOMAIN 24 95 HYDROPHOBIC.
SQ SEQUENCE 183 AA; 18149 MW; 198A5D3B6DF3045A CRC64;

Query Match 64.2%; Score 70; DB 1; Length 183;
Best Local Similarity 78.3%; Pred. No. 0.33;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AAAAAAEEAAAAAEEAAAAA 25
DB 151 AAPAAPAPAAEAPAAEAPAA 173

RESULT 5

ID_HMEL_MOUSE STANDARD; PRT; 401 AA.
AC P09065;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein engrailed-1 (Mo-En-1).
GN EN1 OR EN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93185339; PubMed=1363401;
RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,
RA Provart N.J., Joyner A.L.;
RT "Cloning and sequence comparison of the mouse, human, and chicken
engrailed genes reveal potential functional domains and regulatory
regions.";
RL Dev. Genet. 13:345-358(1992).
RN [2]
RP SEQUENCE OF 278-401 FROM N.A.
RX MEDLINE=88112776; PubMed=2892757;
RA Joyner A.L., Martin G.R.;
RT "En-1 and En-2, two mouse genes with sequence homology to the
Drosophila engrailed gene: expression during embryogenesis.";
RL Genes Dev. 1:29-38(1987).
RN [3]
RP SEQUENCE OF 298-401 FROM N.A.
RX MEDLINE=86079501; PubMed=2416459;
RA Joyner A.L., Kornberg T., Coleman K.G., Cox D.R., Martin G.R.;
RT "Expression during embryogenesis of a mouse gene with sequence
homology to the Drosophila engrailed gene.";
RL Cell 43:29-37(1985).
RN [4]
RP SEQUENCE OF 321-380 FROM N.A.
RX MEDLINE=91099509; PubMed=1980115;
RA Holland P.W.H., Williams N.A.;
RT "Conservation of engrailed-like homeobox sequences during vertebrate
evolution.";
RL FEBS Lett. 277:250-252(1990).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ENGRAILED HOMEBOX FAMILY.
CC -1- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L12703; AAA03660.2; -


```

DR EMBL; Y00201; CAA68361.1; -.
DR PIR; A48423; A48423.
DR HSSP; P02836; 3HDD.
DR TRANSFAC; T02016; -.
DR MGD; MGI:95389; En1.
DR InterPro; IPR000747; Engrailed.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00026; ENGRAILED.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESS.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
DR PROSITE; PS00033; ENGRAILED; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DOMAIN 52 87 PRO-RICH.
FT DOMAIN 73 87 POLY-PRO.
FT DOMAIN 207 228 POLY-ALA.
FT DNA BIND 312 371 HOMEBOX.
SQ SEQUENCE 401 AA; 40950 MW; 1F90210950152FAE CRC64;

Query Match 64.2%; Score 70; DB 1; Length 401;
Best Local Similarity 78.3%; Pred. No. 0.62;
Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAAEAAAAAAAAAAAAA 25
DB 207 AAAAAAAAAAAAAAAAAA 229

RESULT 6
ANP LIMFE STANDARD; PRT; 97 AA.
AC P09031;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Antifreeze protein precursor (AFP).
OS Limanda ferruginea (Yellowtail flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Limanda.
OX NCBI_TaxID=8258;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88029483; PubMed=3665937;
RA Scott G.K., Davies P.L., Shears M.A., Fletcher G.L.;
RT "Structural variations in the alanine-rich antifreeze proteins of the pleuronectinae."
RL Eur. J. Biochem. 168:629-633 (1987).
CC 1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC 1- SIMILARITY: BELONGS TO THE TYPE-1 AFP FAMILY. TYPE 1 AFP ARE ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; X06356; CAA29655.1; -.
DR PIR; S02376; S02376.
DR InterPro; IPR000104; Antifreeze 1.
DR PRINTS; PR00308; ANTIFREEZE1.
KW Antifreeze protein; Repeat; Signal.
FT SIGNAL 1 23
FT PROPEP 24 48 REMOVED BY A DIPEPTIDYLPEPTIDASE

FT CHAIN 49 97 (PROBABLE).
SQ SEQUENCE 97 AA; 8865 MW; 62AD582DF8E459B6 CRC64;
ANTIFREEZE PROTEIN.
62AD582DF8E459B6 CRC64;

Query Match 63.3%; Score 69; DB 1; Length 97;
Best Local Similarity 70.8%; Pred. No. 0.26;
Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAEAAAAAAAAAAAAA 25
DB 53 DAAAAATAAATAAAKAADTAAA 76

RESULT 7
OVO DROME STANDARD; PRT; 1028 AA.
AC P51521; O9X2UA;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE OVO protein (Shaven baby protein).
GN OVO OR SVB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=95021209; PubMed=7935398;
RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster: relationship to genetic complexity."
RL Mol. Cell. Biol. 14:6809-6818 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=91293102; PubMed=1712294;
RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
RT "The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development."
RL EMO J. 10:2259-2266 (1991).
CC 1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC 1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC 1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG, BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC 1- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; U11383; AAB60216.1; -.
DR EMBL; X59772; CAB36921.1; ALT_SEQ.
DR PIR; A56038; A56038.
DR HSSP; P07248; 2ADR.
DR TRANSFAC; T00669; -.
DR FlyBase; FBgn0003028; ovo.
DR InterPro; IPR007087; Znf C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR SMART; SM00355; Znf C2H2; 4.
DR PROSITE; PS00028; ZINC FINGER C2H2_1; 3.
DR PROSITE; PS01571; ZINC FINGER C2H2_2; 3.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein; Transcription regulation.

```

```

FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-GLY.
FT DOMAIN 144 152 POLY-HIS.
FT DOMAIN 153 159 POLY-ASN.
FT DOMAIN 336 339 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 549 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 823 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT ZN_FING 874 896 C2H2-TYPE 1.
FT ZN_FING 902 924 C2H2-TYPE 2.
FT ZN_FING 930 953 C2H2-TYPE 3.
FT ZN_FING 969 992 C2H2-TYPE 4.
FT ZN_FING 992 992 C2H2-TYPE 4.
FT CONFLICT 647 647 A > R (IN REF. 2).
SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0F6F77 CRC64;

Query Match 63.3%; Score 69; DB 1; Length 1028;
Best Local Similarity 85.7%; Pred. No. 1.6;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAEEEEEEEEAAAAA 23
Db 497 AAAAAAEEEEEEEEAAAAA 517

RESULT 8
ASX DROME STANDARD; PRT; 1669 AA.
AC QSV727: 076930;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polycomb protein Asx (Additional sex combs).
GN ASX OR CG8787.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND DEVELOPMENTAL
RP STAGE.
RC TISSUE=Imaginal disks;
RX MEDLINE=98146384; PubMed=9477319;
RA Sinclair D.A.R., Milne T.A., Hodgson J.W., Shellard J., Salinas C.A.,
RA Kyba M., Randazzo F., Brock H.W.;
RT "The Additional sex combs gene of Drosophila encodes a chromatin
RT protein that binds to shared and unique Polycomb group sites on
RT polytene chromosomes.";
RL Development 125:1207-1216 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,

```

```

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Franckoch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman D.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J.P., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Rector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [3]
RP INTERACTION WITH TAN.
RX MEDLINE=21290825; PubMed=11397012;
RA Dietrich B.H., Moore J., Kyba M., dosSantos G., McCloskey F.,
RA Milne T.A., Brock H.W., Krause H.M.;
RT "Tantalus, a novel ASX-interacting protein with tissue-specific
RT functions.";
RL Dev. Biol. 234:441-453 (2001).
CC -!- FUNCTION: Atypical Polycomb group protein, which may be involved
CC in both Polycomb group (PcG) and trithorax group (trxG) complexes.
CC PcG and trxG proteins act by forming multiprotein complexes, which
CC are respectively required to maintain the transcriptionally
CC repressive and transcriptionally active state of homeotic genes
CC throughout development. PcG and trxG protein complexes are not
CC required to initiate repression and activation, but to maintain it
CC during later stages of development. Both complexes probably act
CC via methylation of histones, rendering chromatin heritably changed
CC in its expressibility.
CC -!- SUBUNIT: Interacts with Tan.
CC -!- SUBCELLULAR LOCATION: Nuclear; associated with chromatin.
CC Colocalizes with many PcG sites on polytene chromosomes.
CC -!- TISSUE SPECIFICITY: Highly expressed in nurse cells and deposited
CC in oocytes late in oogenesis. Ubiquitous in early embryos. Late
CC embryos show higher levels in CNS and neuroectoderm.
CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC Early embryos have high levels of expression, this drops off and
CC zygotic expression begins at 3-6 hour embryos. Expression levels
CC are low in larvae and medium in pupae and adults.
CC -!- DOMAIN: Contains two Leu-Xaa-Xaa-Leu (LXXLL) motifs, which may
CC be required for an association with nuclear receptors (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Asx family.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to
CC frameshifts in positions 608 and 719.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AJ001164; CAA04568.1; ALT_FRAME.
 DR EMBL; AB003814; RAF58239.1; -
 DR FlyBase; FBgn000142; ABX.
 DR PROSITE; PS01359; ZF_PHD_1; FALSE NEG.
 DR PROSITE; PS0016; ZF_PHD_2; FALSE NEG.
 KW Transcription regulation; Repressor; Nuclear protein; Zinc;
 KW Metal-binding; Zinc-finger; Repeat; Developmental protein.
 FT ZN FING 1632 1669 PHD-TYPE (ATYPICAL).
 FT DOMAIN 8 12 POLY-THR.
 FT DOMAIN 122 126 POLY-GLN.
 FT DOMAIN 129 152 ALA-RICH.
 FT DOMAIN 638 715 SER-RICH.
 FT DOMAIN 747 751 POLY-GLN.
 FT DOMAIN 862 1202 GLN-RICH.
 FT DOMAIN 1287 1290 POLY-THR.
 FT DOMAIN 1520 1524 POLY-HIS.
 FT DOMAIN 1527 1536 POLY-GLN.
 FT SITE 224 228 LXXLL MOTIF 1.
 FT SITE 244 248 LXXLL MOTIF 2.
 FT CONFLICT 14 15 SQ -> CE (IN REF. 1).
 FT CONFLICT 187 187 K -> N (IN REF. 1).
 FT CONFLICT 183 1253 S -> T (IN REF. 1).
 FT CONFLICT 1520 1520 MISSING (IN REF. 1).
 SQ SEQUENCE 1669 AA; 179841 MW; F65D87398D67D321 CRC64;

Query Match 63.3%; Score 69; DB 1; Length 1669;
 Best Local Similarity 78.3%; Pred. No. 2.4;
 Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAATAAATAAATAAATAAATAA 24
 :|||||
 Db 128 KAATAAATAAATAAATAAATAAQA 150

RESULT 9

SIM DROME
 ID SIM DROME STANDARD; PRT; 697 AA.
 AC P05709; O96521; Q8M017; Q9VEZ3;
 DT 01-NOV-1998 (Rel. 09, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Single-minded protein.
 GN SIM OR CG7771.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 EX MEDLINE=990545; PubMed=9840810;
 RA Kaeali Y., Stahl S., Crews S.;
 RT "Specification of the Drosophila CNS midline cell lineage: direct
 RT control of single-minded transcription by dorsal/ventral patterning
 RT genes.";
 RL Gene Expr. 7:171-189 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Adams M.D., Celniker S.E., Richards S., Ashburner M., Henderson S.N.,
 RA George R.A., Lewis S.E., Yandell M.D., Zhang Q., Chen L.X.,
 RA Sutton G.G., Wortman J.R., White M.J., Blakes R.G., Chang M., Pfeiffer B.D.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer J.C., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burkis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Foler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glöde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostali D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jatalin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Jimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Ial Z.,
 RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN [3]
 RP REVISIONS.
 RC STRAIN=Berkeley;
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A Drosophila full-length cDNA resource.";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
 RN [5]
 RP SEQUENCE OF 25-42 FROM N.A., AND SIMILARITY TO HLH PROTEINS.
 RX MEDLINE=92103681; PubMed=1760843;
 RA Nambu J.R., Lewis J.O., Wharton K.A. Jr., Crews S.T.;
 RT "The Drosophila single-minded gene encodes a helix-loop-helix protein
 RT that acts as a master regulator of CNS midline development.";
 RL Cell 67:1157-1167 (1991).
 RN [6]
 RP SEQUENCE OF 43-597 FROM N.A.
 RX MEDLINE=88151023; PubMed=3345560;
 RA Crews S.T., Thomas J.B., Goodman C.S.;
 RT "The Drosophila single-minded gene encodes a nuclear protein with
 RT sequence similarity to the per gene product.";
 RL Cell 52:143-151 (1988).
 CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT FUNCTIONS AS A MASTER
 CC DEVELOPMENTAL REGULATOR OF THE CNS MIDLINE LINEAGE. MUTATIONS IN
 CC THE SIM GENE RESULTS IN THE LOSS OF THE PRECURSOR CELLS GIVING
 CC RISE TO MIDLINE CELLS OF THE EMBRYONIC CENTRAL NERVOUS SYSTEM.
 CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another
 CC HLH protein.

CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- POLYMORPHISM: Berkeley strain has 11 A-A-Q repeats.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS.
 CC -1- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
 CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF071934; AAC64519.1; ALT_SEQ.
 CC DR EMBL: AE003698; AAF54902.3; -.
 CC DR EMBL: AY129457; AAU76199.1; -.
 CC DR EMBL: M19020; AAA28900.1; -.
 CC DR F01: A29945; A29945.
 CC DR TRANSFAC: T00750; -.
 CC DR FlyBase: FBgn0004666; sim.
 CC DR GO: GO:0005634; C:nucleus; IEP.
 CC DR GO: GO:0003702; F:RNA polymerase II transcription factor acti...; NAS.
 CC DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC DR GO: GO:0007418; P:ventral midline development; IMP.
 CC DR InterPro: IPR001092; HLH basic.
 CC DR InterPro: IPR001067; Nuc_translocat.
 CC DR InterPro: IPR001610; PAC.
 CC DR InterPro: IPR000014; PAC_domain.
 CC DR Pfam: PF00010; HLH; 1.
 CC DR Pfam: PF00785; PAC; 1.
 CC DR Pfam: PF00989; PAS; 1.
 CC DR PRINTS: PR00785; NCTNSLOCATR.
 CC DR SMART: SM00353; HLH; 1.
 CC DR SMART: SM00086; PAC; 1.
 CC DR SMART: SM00091; PAS; 2.
 CC DR TIGRFAMs: TIGR00229; sensory_box; 2.
 CC DR PROSITE: PS00038; HLH 1; 1.
 CC DR PROSITE: PS00888; HLH 2; 1.
 CC DR PROSITE: PS0112; PAS; 2.
 CC Developmental protein; Neurogenesis; Nuclear protein; Repeat;
 KW Transcription regulation; DNA-binding.
 FT DNA_BIND 21 37
 FT DOMAIN 38 78
 FT DOMAIN 100 172
 FT DOMAIN 266 336
 FT DOMAIN 406 446
 FT DOMAIN 553 672
 FT DOMAIN 673 693
 FT VARIANT 406 414 MISSING (IN STRAIN BERKELEY).
 FT CONFLICT 151 151 I -> Y (IN REF. 1).
 FT SEQUENCE 697 AA; 76475 MW; 58841444A17101AD CRC64;
 SQ
 Query Match 62.4%; Score 68; DB 1; Length 697;
 Best Local Similarity 66.7%; Pred. No. 1.5;
 Matches 16; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 2 EAARAAAEAAAEAAAEAAAEAAAEAA 25
 DB 408 QAAQAQAQAQAQAQAQAQAQAQA 431
 RESULT 10
 ID HR38_DROME STANDARD; PRT; 1073 AA.
 AC P49869; O18383; Q9V1K4;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable nuclear hormone receptor HR38 (GHR38).

GN HR38 OR NR4A4 OR CG1864.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).
 RC TISSUE=Larva;
 RX MEDLINE=95372400; PubMed=7644522;
 RA Sutherland J.D., Kozlova T., Tzerzinis G., Kafatos F.C.;
 RT "Drosophila hormone receptor 38: a second partner for Drosophila USP
 RT suggests an unexpected role for nuclear receptors of the nerve growth
 RT factor-induced protein B type.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7966-7970(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND TISSUE SPECIFICITY.
 RX MEDLINE=98370123; PubMed=9704500;
 RA Komonyi O., Mink M., Csiba J., Maroy P.;
 RT "Genomic organization of DHR38 gene in Drosophila: presence of
 RT Alu-like repeat in a translated exon and expression during embryonic
 RT development.";
 RL Arch. Insect Biochem. Physiol. 38:185-192(1998).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM LONG), AND CHARACTERIZATION.
 RC TISSUE=Larva;
 RX MEDLINE=98315108; PubMed=9649534;
 RA Kozlova T., Pokholkova G.V., Tzerzinis G., Sutherland J.D.,
 RA Zhimulev I.F., Kafatos F.C.;
 RT "Drosophila hormone receptor 38 functions in metamorphosis: a role in
 RT adult cuticle formation.";
 RL Genetics 149:1465-1475(1998).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM LONG).
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton R.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 Wan K.H., Doyle C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Franckoch C., Baldwin D.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hoskins D., Houston K.A., Howland T.J., Wei M.-H., Ibegawa C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Mosnrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195 (2000).
RN [5]
RP SEQUENCE OF 528-1073 FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=96068664; PubMed=7479849;
RA Fiek G.J., Thummel C.S.;
RT Isolation, regulation, and DNA-binding properties of three
RT Drosophila nuclear hormone receptor superfamily members";
RL Proc. Natl. Acad. Sci. U.S.A. 92:10604-10608 (1995).
CC -1- FUNCTION: BINDS TO NGFI-B RESPONSE ELEMENTS. PLAYS AN IMPORTANT
CC ROLE IN LATE STAGES OF EPIDERMAL METAMORPHOSIS.
CC -1- SUBUNIT: FORMS A HETERODIMER WITH USP.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P49869-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P49869-2; Sequence=VSP_003714;
CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN PREBLASTODERM
CC EMBRYOS, SPECIFICALLY IN CENTRAL NERVOUS SYSTEM AND INTESTINAL
CC TRACT. HIGHLY EXPRESSED IN THIRD INSTAR LARVAL IMAGINAL DISKS AND
CC BRAIN COMPLEXES, BUT NOT IN OVARIES.
CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN 0-8 HOUR EMBRYOS AND ADULTS.
CC HIGHER IN LATE EMBRYOGENESIS AND DURING LARVAL AND PUPAL STAGES.
CC SHORT ISOFORM IS ENRICHED IN PUPAE AND ADULTS, LONG ISOFORM IN
CC LARVAE.
CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X89246; CAA61534.1; -
DR EMBL; Y15606; CAA75690.1; -
DR EMBL; AJ002073; CAA05172.1; -
DR EMBL; AE003667; AAF53914.1; -
DR EMBL; U36762; AAC46926.1; -
DR HSPF; P19793; 2NLL.
DR TRANSFAC; T02760; -
DR FlyBase; FBgn0014859; Hx38.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrm_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR PRODOM; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI_1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Alternative splicing; Developmental protein.
FT DNA_BIND 744 809
FT ZN_FING 744 809
FT ZN_FING 744 804
FT ZN_FING 780 804
FT DOMAIN 188 192
FT DOMAIN 206 218
FT DOMAIN 221 228
FT DOMAIN 268 272
FT DOMAIN 294 312
FT DOMAIN 441 462
FT DOMAIN 505 508
FT DOMAIN 619 626
FT DOMAIN 661 665
FT VARSPPLIC 1 522
Missing (in isoform Short).
/FTId=VSP_003714.

FT CONFLICT 667 667 V -> VSSPSV (IN REF. 4).
FT CONFLICT 685 685 S -> L (IN REF. 1 AND 3).
FT CONFLICT 689 692 STAQ -> LEGER (IN REF. 2).
FT CONFLICT 697 697 A -> D (IN REF. 2).
FT CONFLICT 702 702 N -> S (IN REF. 2).
FT CONFLICT 1041 1041 S -> R (IN REF. 2).
FT CONFLICT 1064 1064 E -> D (IN REF. 2).
SQ SEQUENCE 1073 AA; 116991 MW; 126A30DAFAIC096A CRC64;

Query Match 62.4%; Score 68; DB 1; Length 1073;
Best Local Similarity 73.9%; Pred. No. 2.1;
Matches 17; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 AAAAAAAAAAAAAAAAAAAAAA 25
Db 206 AATAAATAAAGGAAASAAAA 228

RESULT 11
PMXB HUMAN
ID PMXB HUMAN STANDARD; PRT; 314 AA.
AC Q99453;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)
DE (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).
GN PHOX2B OR PMXB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=97191543; PubMed=9039501;
RA Yokoyama M., Nishi Y., Yoshii J., Okubo K., Matsubara K.;
RT "Identification and cloning of neuroblastoma-specific and nerve
RT tissue-specific genes through compiled expression profiles";
RL DNA Res. 3:311-320 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20487360; PubMed=11034547;
RA Adachi M., Browne D., Lewis E.J.;
RT "Paired-like homeodomain proteins Phox2a/Arx and Phox2b/NBPhox have
RT similar genetic organization and independently regulate dopamine
RT beta-hydroxylase gene transcription";
RL DNA Cell Biol. 19:539-554 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99326521; PubMed=10395798;
RA Yokoyama M., Watanabe H., Nakamura M.;
RT "Genomic structure and functional characterization of NBPhox (PMXB2B),
RT a homeodomain protein specific to catecholaminergic cells that is
RT involved in second messenger-mediated transcriptional activation";
RL Genomics 59:40-50 (1999).
CC -1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF SEVERAL MAJOR
CC NORADRENERGIC NEURON POPULATIONS, INCLUDING THE LOCUS COERULEUS.
CC TRANSCRIPTION FACTOR WHICH COULD DETERMINE A NEUROTRANSMITTER
CC PHENOTYPE IN VERTEBRATES. ENHANCES SECOND-MESSENGER-MEDIATED
CC ACTIVATION OF THE DOPAMINE BETA-HYDROXYLASE AND C-FOS PROMOTERS,
CC AND OF SEVERAL ENHANCERS INCLUDING CYCLIC AMP-RESPONSE ELEMENT AND
CC SERUM-RESPONSE ELEMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEUROBLASTOMA, BRAIN AND ADRENAL
CC GLAND.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; D82344; BAA11555.1; --
 DR EMBL; AF117979; AAD26698.1; --
 DR EMBL; AB015671; BAA82670.1; --
 DR PIR; JCS273; JCS273.
 DR HSSP; P06601; IFJUL.
 DR TRANSFAC; T03961; --
 DR Genew; HGNC:9143; PHOX2B.
 DR MIM; 603851; --
 DR GO; GO:0003712; F:transcription cofactor activity; TAS.
 DR GO; GO:0003700; F:transcription factor activity; TAS.
 DR GO; GO:0007399; P:neurogenesis; TAS.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS00071; HOMEBOX 2; 1.
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation; Activator.
 FT DNA BIND 98 157 HOMEBOX.
 FT DOMAIN 159 167 POLY-ALA.
 FT DOMAIN 212 217 POLY-GLY.
 FT DOMAIN 241 260 POLY-ALA.
 SQ SEQUENCE 314 AA; 31607 MW; 76737F71948B5D81 CRC64;

Query Match 61.5%; Score 67; DB 1; Length 314;
 Best Local Similarity 78.3%; Pred. No. 1;
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AAAAAEAAAAAAAAAAAAA 25
 ||||| ||||| ||||| |||||
 Db 244 AAAAAEAAAAAAAAAAAAA 266

RESULT 12

ID PMXB_MOUSE STANDARD; PRT; 314 AA.
 AC Q35690;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)
 DE (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).
 GN PHOX2B OR PMXB2B
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98040559; PubMed=9374403;
 RA Pattyn A., Morin X., Cremer H., Goridis C., Brunet J.-F.;
 RT "Expression and interactions of the two closely related homeobox
 genes Phox2a and Phox2b during neurogenesis.";
 RL Development 124:4065-4075(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99326521; PubMed=10395798;
 RA Yokoyama M., Watanabe H., Nakamura M.;
 RT "Genomic structure and functional characterization of NBPhox (PMXB2B),
 a homeodomain protein specific to catecholaminergic cells that is
 involved in second messenger-mediated transcriptional activation.";
 RL Genomics 59:40-50(1999).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; Y14493; CAA74833.1; --
 DR HSSP; P06601; IFJUL.
 DR TRANSFAC; T03976; --
 DR MGD; MGI:1100882; Phox2b.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS00071; HOMEBOX 2; 1.
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 KW DNA BIND 98 157 HOMEBOX.
 FT DOMAIN 159 167 POLY-ALA.
 FT DOMAIN 212 217 POLY-GLY.
 FT DOMAIN 241 260 POLY-ALA.
 SQ SEQUENCE 314 AA; 31621 MW; 40737F71948B595A CRC64;

Query Match 61.5%; Score 67; DB 1; Length 314;
 Best Local Similarity 78.3%; Pred. No. 1;
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AAAAAEAAAAAAAAAAAAA 25
 ||||| ||||| ||||| |||||
 Db 244 AAAAAEAAAAAAAAAAAAA 266

RESULT 13

ID MAZ_HUMAN STANDARD; PRT; 477 AA.
 AC P56270; Q15703; Q99443;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Myc-associated zinc finger protein (MAZI) (Purine-binding
 DE transcription factor) (Pur-1) (ZFP87) (ZIF87).
 GN MAZ.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92366479; PubMed=1502157;
 RA Bossone S.A., Asselin C., Patel A.J., Marcu K.B.;
 RT "MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences
 RT regulating transcriptional initiation and termination.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7452-7456(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Carcinoma;
 RX MEDLINE=92232709; PubMed=1567856;
 RA Pyrc J.J., Moberg K.H., Hall D.J.;
 RT "Isolation of a novel cDNA encoding a zinc-finger protein that binds
 RT to two sites within the c-myc promoter.";
 RL Biochemistry 31:4102-4110(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Pancratic islets;
 RX MEDLINE=96428591; PubMed=8831693;
 RA Teutsui H., Sakatsume O., Itakura K., Yokoyama K.K.;
 RT "Members of the MAZ family: a novel cDNA clone for MAZ from human
 RT pancreatic islet cells";
 RL Biochem. Biophys. Res. Commun. 226:801-809(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96224025; PubMed=8626793;

RA Parks C.L., Shenk T.;
 RT "The serotonin 1a receptor gene contains a TATA-less promoter that
 responds to MAZ and Sp1.";
 RL J. Biol. Chem. 271:4417-4430 (1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphoblastoma;
 RX MEDLINE=98352105; PubMed=9685418;
 RA Song J., Murakami H., Teutsui H., Tang X., Matsumura M., Itakura K.,
 RA Kanazawa I., Sun K., Yokoyama K.K.;
 RT "Genomic organization and expression of a human gene for Myc-
 associated zinc finger protein (MAZ).";
 RL J. Biol. Chem. 273:20603-20614 (1998).
 CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR WITH DUAL ROLES
 CC IN TRANSCRIPTION INITIATION AND TERMINATION. BINDS TO TWO SITES,
 CC MEI1 AND MEI2, WITHIN THE C-MYC PROMOTER HAVING GREATER
 CC AFFINITY FOR THE FORMER. ALSO BINDS TO MULTIPLE G/C-RICH SITES
 CC WITHIN THE PROMOTER OF THE SPI FAMILY OF TRANSCRIPTION FACTORS.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL
 CC MUSCLE, AND PANCREAS. SEEMS NOT TO BE EXPRESSED IN KIDNEY.
 CC -!- SIMILARITY: Contains 6 C2H2-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M94046; -; NOT ANNOTATED CDS.
 CC EMBL; M93339; -; NOT ANNOTATED CDS.
 CC EMBL; D85131; BA12728.1; ALT INIT.
 CC EMBL; U33819; AAB04121.1; ALT INIT.
 CC EMBL; AB017335; BAA33064.1; -;
 CC PIR; A42170; A42170.
 CC TRANSFAC; T00490; -;
 CC TRANSFAC; T02305; -;
 CC Genew; HGNC:6914; MAZ.
 CC MIM; 600999; -;
 CC GO; GO:0006367; P:transcription initiation from Pol II promoter; TAS.
 CC GO; GO:0006369; P:transcription termination from Pol II promoter; TAS.
 CC InterPro; IPR007087; Znf_C2H2.
 CC Pfam; PF00096; ZF-C2H2; 6.
 CC ProDom; PD000003; Znf_C2H2; 1.
 CC SMART; SM00355; Znf_C2H2; 6.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_2; 5.
 CC PROSITE; PS0157; ZINC_FINGER_C2H2_1; 5.
 CC Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KW RNA-binding; Repeat; Nuclear protein.
 FT ZN_FING 190 212
 FT C2H2-TYPE 1.
 FT ZN_FING 279 301
 FT C2H2-TYPE 2.
 FT ZN_FING 307 329
 FT C2H2-TYPE 3.
 FT ZN_FING 337 360
 FT C2H2-TYPE 4.
 FT ZN_FING 366 388
 FT C2H2-TYPE 5.
 FT ZN_FING 392 413
 FT C2H2-TYPE 6.
 FT DOMAIN 96 108
 FT POLY-ALA.
 FT DOMAIN 133 139
 FT POLY-PRO.
 FT DOMAIN 157 161
 FT POLY-ALA.
 FT DOMAIN 245 249
 FT POLY-GLY.
 FT DOMAIN 435 449
 FT POLY-ALA.
 FT CONFLICT 259 259
 FT MISSING (IN REF. 3).
 FT CONFLICT 401 401
 FT L -> M (IN REF. 2 AND 4).
 FT CONFLICT 443 447
 FT MISSING (IN REF. 3).
 SQ SEQUENCE 477 AA; 48607 MW; C04C80F32C3C6825 CRC64;
 Query Match 61.5%; Score 67; DB 1; Length 477;
 Best Local Similarity 77.3%; Pred. No. 1.4;
 Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 4 AAAAAAAAEAAAAAAAAAAAAA 25
 ||| :||| ||||| ||||| |||||

Db 90 AAOESAAAAAAAAAAAAAAAAA 111
 RESULT 14
 HM1D DROAN
 ID HM1D DROAN STANDARD; PRT; 606 AA.
 AC P22544;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Homeobox protein B-H1.
 GN B-H1 OR OM (LD).
 OS Drosophila ananassae (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CC NCBI_TaxID=7217;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91122048; PubMed=1671353;
 RA Tanda S., Corces V.G.;
 RT "Retrotroposon-induced overexpression of a homeobox gene causes
 defects in eye morphogenesis in Drosophila.";
 RL EMBO J. 10:407-417 (1991).
 CC -!- FUNCTION: Probably involved in eye morphogenesis.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X56682; CAA40011.1; -;
 CC PIR; S13367; S13367.
 CC HSSP; P14653; 1B72.
 CC TRANSFAC; T03732; -;
 CC FlyBase; FBgn012114; Dana\B-H1.
 CC InterPro; IPR001356; Homeobox.
 CC Pfam; PF00046; homeobox; 1.
 CC PRINTS; PR00024; HOMEOBOX.
 CC ProDom; PD000010; Homeobox; 1.
 CC SMART; SM00389; HOX; 1.
 CC PROSITE; PS00027; HOMEOBOX_1; 1.
 CC PROSITE; PS00071; HOMEOBOX_2; 1.
 CC DNA-binding; Homeobox; Developmental protein; Nuclear protein; Vision.
 FT DOMAIN 23 57
 FT HIS/GLN-RICH (OPA-REPEAT).
 FT DOMAIN 106 124
 FT HIS/GLN-RICH (OPA-REPEAT).
 FT DOMAIN 173 193
 FT HIS/PRO-RICH.
 FT DOMAIN 331 390
 FT HOMEOBOX.
 FT DOMAIN 220 248
 FT ALA-RICH.
 FT DOMAIN 422 434
 FT ALA-RICH.
 FT DOMAIN 450 455
 FT ALA-RICH.
 FT DOMAIN 503 510
 FT ALA-RICH.
 FT DOMAIN 515 521
 FT PRO-RICH.
 SQ SEQUENCE 606 AA; 61735 MW; AA7B8B6367370PBB CRC64;
 Query Match 61.5%; Score 67; DB 1; Length 606;
 Best Local Similarity 78.3%; Pred. No. 1.7;
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 3 AAAAAAAAEAAAAAAAAAAAAA 25
 ||| :||| ||||| ||||| |||||

Db 220 AAAAAAAARNNSAAAAAAAAAAAA 242
 RESULT 15
 HXAD MOUSE
 ID HXAD MOUSE STANDARD; PRT; 386 AA.
 AC Q62424;

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Homeobox protein Hox-A13 (Hox-1.10).
 GN HOXA13 OR HOX-1.10.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=96259555; PubMed=86731126;
 RX Mortlock D.P., Post L.C., Innis J.W.;
 RA "The molecular basis of Hypodactyly (Hd): a deletion in Hoxa 13 leads
 RT to arrest of digital arch formation.";
 RL Nat. Genet. 13:284-289(1996).
 CC -|- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS (BY
 CC SIMILARITY).
 CC -|- SUBCELLULAR LOCATION: Nuclear.
 CC -|- DISEASE: DEFECTS IN HOXA13 ARE THE CAUSE OF HYPODACTYLY (HD), A
 CC CONDITION CHARACTERIZED BY PROFOUND DEFICIENCY OF DIGITAL ARCH
 CC STRUCTURES.
 CC -|- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U59322; AAB03322.1; -
 CC HSSP; P14653; 1B72.
 CC TRANSFAC; T03337; -
 CC MGD; MGI:96173; Hoxa13.
 CC InterPro; IPR001356; Homeobox.
 CC Pfam; PF00046; homeobox; 1.
 CC ProDom; PD000010; Homeobox; 1.
 CC SMART; SM00389; HOX; 1.
 CC PROSITE; PS00027; HOMEBOX 1; 1.
 CC PROSITE; PS00071; HOMEBOX_2; 1.
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation.
 FT DOMAIN 38 51 POLY-ALA.
 FT DNA_BIND 320 379 HOMEBOX.
 FT DOMAIN 52 57 POLY-GLY.
 FT DOMAIN 62 66 POLY-ALA.
 FT DOMAIN 73 84 POLY-ALA.
 FT DOMAIN 101 104 POLY-ALA.
 FT DOMAIN 116 133 POLY-ALA.
 FT DOMAIN 198 205 POLY-ALA.
 SQ SEQUENCE 386 AA; 39566 MW; 2B01DCC9B1951324 CRC64;

Query Match 60.6%; Score 66; DB 1; Length 386;
 Best Local Similarity 70.8%; Pred. No. 1.5;
 Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAEEEEAAAAAEEAAAAA 25
 || : ||| ||| ||| ||| :
 Db 111 EAPPSSAAAAAEEEEAAAAAAS 134

Search completed: January 30, 2004, 00:20:45
 Job time : 6.6338 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:52 ; Search time 25.8216 Seconds
(without alignments)
249.842 Million cell updates/sec

Title: US-09-461-684C-2
Perfect score: 109
Sequence: 1 CCAAAAEAAAAAAAAAAAAA 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archesp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	84	77.1	1038	5	Q8MQW9	Q8mqw9 drosophila
2	84	77.1	2347	5	Q8INH9	Q8inh9 drosophila
3	84	77.1	2451	5	Q9VG05	Q9vg05 drosophila
4	82	75.2	378	11	Q8R089	Q8r089 mus musculu
5	80	73.4	301	11	Q8BJK2	Q8bjk2 mus musculu
6	80	73.4	1354	11	Q9EPW8	Q9epw8 mus musculu
7	79	72.5	110	11	Q9LW00	Q9lw00 mus musculu
8	79	72.5	332	4	Q8IZU1	Q8izu1 homo sapien
9	77	70.6	265	10	Q93598	Q93598 chlamydomon
10	76	69.7	665	11	Q9QXG2	Q9qxg2 mus musculu
11	76	69.7	1452	4	Q9H4A0	Q9h4a0 homo sapien
12	76	69.7	1512	4	Q9H4A1	Q9h4a1 homo sapien
13	75	68.8	221	5	Q9VXD3	Q9vxd3 drosophila
14	72	66.1	997	5	Q9W2J2	Q9w2j2 drosophila
15	71	65.1	206	10	Q93597	Q93597 chlamydomon
16	71	65.1	1216	10	Q95WH3	Q95wh3 chlamydomon

17	71	65.1	1787	10	Q9M4X9	Q9mix9 chlamydomon
18	70	64.2	246	16	Q8G3U2	Q8g3u2 bifidobacte
19	70	64.2	323	4	Q9H782	Q9h782 homo sapien
20	70	64.2	349	4	O43856	O43856 homo sapien
21	70	64.2	423	4	O75400	O75400 homo sapien
22	70	64.2	452	4	O75404	O75404 homo sapien
23	70	64.2	484	11	Q923T4	Q923t4 mus musculu
24	70	64.2	512	4	Q8NAP8	Q8nap8 homo sapien
25	70	64.2	618	11	Q8BX04	Q8bx04 mus musculu
26	70	64.2	618	11	Q8BUW9	Q8buw9 mus musculu
27	70	64.2	1171	3	Q9P3E2	Q9p3e2 neurospora
28	69	63.3	218	11	Q9CS14	Q9csi4 mus musculu
29	69	63.3	512	11	Q8CB12	Q8cb12 mus musculu
30	69	63.3	1222	5	Q8T8L9	Q8t8l9 drosophila
31	69	63.3	1354	5	Q8MPN4	Q8mpn4 drosophila
32	69	63.3	1668	5	O76930	O76930 drosophila
33	69	63.3	1669	5	Q9V727	Q9v727 drosophila
34	68	62.4	403	16	Q9JUT2	Q9jut2 neisseria m
35	68	62.4	543	5	Q9W4F9	Q9w4f9 drosophila
36	68	62.4	545	12	Q91TR2	Q91tr2 tupaia herp
37	68	62.4	1340	16	Q9L1H8	Q9l1h8 streptomyc
38	68	62.4	2023	5	Q9V5Z9	Q9v5z9 drosophila
39	68	62.4	2023	5	O96542	O96542 drosophila
40	67.5	61.9	324	5	Q9NGI7	Q9ngi7 drosophila
41	67.5	61.9	324	5	Q9NGI8	Q9ngi8 drosophila
42	67.5	61.9	324	5	Q9NGI9	Q9ngi9 drosophila
43	67.5	61.9	324	5	Q9NGI9	Q9ngi9 drosophila
44	67.5	61.9	324	5	Q9NGI9	Q9ngi9 drosophila
45	67	61.5	324	5	Q9NGB4	Q9ngb4 drosophila

ALIGNMENTS

RESULT 1

Q8MQW9 PRELIMINARY; PRT; 1038 AA.

AC Q8MQW9;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-WAR-2003 (TREMBLrel. 23, Last annotation update)
DE SD05989p (Fragment).
GN CG7518.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY122252; AAM52764.1; -;
DR FlyBase; FBgn0038108; CG7518.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PRO0308; ANTIFREEZE1.
DR PRINTS; PRO1217; PRICEXTENS.
FT NON TER 1
SQ SEQUENCE 1038 AA; 109059 MW; 80C935A2C6DBA276 CRC64;

Query Match 77.1%; Score 84; DB 5; Length 1038;

Best Local Similarity 91.3%; Pred.No. 1.1;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AAAAAAAAAAAAAAAAAAAAAA 25

Db 112 AAAAAAAAAAAAAAAAAAAAAA 134

```

RESULT 2
ID Q8INH9 PRELIMINARY; PRT: 2347 AA.
AC Q8INH9;
DT 01-WAR-2003 (TRENBLrel. 23, Created)
DT 01-WAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-WAR-2003 (TRENBLrel. 23, Last annotation update)
DE CG7518-PB.
GN CG7518.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Evansgelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL SCIENCE 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celnik S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banson J., An H., Baldwin D., Banson J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
RA Dodson K.J., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.P., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas S., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."

```

Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[3]

SEQUENCE FROM N.A.

RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kamniker J.S., Prochnik S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnik S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield B.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,

PT "Annotation of Drosophila melanogaster genome";

Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[4]

SEQUENCE FROM N.A.

RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;

RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[5]

SEQUENCE FROM N.A.

RA Flybase;

RL Submitted (SBP-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003698; AANI4338.1; -

SQ SEQUENCE 2347 AA; 257013 MW; 23BF5FC5FFCAEA64 CRC64;

Query Match 77.1%; Score 84; DB 5; Length 2347;

Best Local Similarity 91.3%; Pred. No. 2.3;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAAAAAAAAAAAAA 25

DB 1374 AAAAAAAAAAAAAAAAAAAA 1396

RESULT 3

Q8VG05 PRELIMINARY; PRT: 2451 AA.

AC Q8VG05;

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE CG7518 protein.

GN CG7518.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y., Helt G., Nelson C.R., Gabor G.L.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K.J., Evansgelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,

RA Glodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,

RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,

RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,

RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL SCIENCE 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Celnik S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banson J., An H., Baldwin D., Banson J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,

RA Dodson K.J., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferreira S., Frise E., Galle R.P., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb J., Paragas S., Park S., Patel S., Pfeiffer B.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RT "Sequencing of Drosophila melanogaster genome.";

RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri F.N., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of *Drosophila melanogaster*;"
RL Science 287:2185-2195(2000).
DR EMBL; AB003698; AAF54888.2; --
DR FlyBase; FBgn0038108; CG7518.
DR InterPro; IPR001005; Myb_DNA_binding.
DR PROSITE; PS00037; MYB_1; 1.
SQ SEQUENCE 2451 AA; 266959 MW; 088A2293F27481E2 CRC64;

Query Match 77.1%; Score 84; DB 5; Length 2451;
Best Local Similarity 91.3%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAAAAAAAAAAAAA 25
|||||
Db 1374 AAAAAAAAAAAAAAAAAAAAAA 1396

RESULT 4

Q8R089 ID Q8R089 PRELIMINARY; PRT; 378 AA.
AC Q8R089;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to hypothetical protein FLJ11618.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
KW Hypothetical protein.
SQ SEQUENCE 378 AA; 39456 MW; 4C3FAF0D4AC29869 CRC64;

Query Match 75.2%; Score 82; DB 11; Length 378;
Best Local Similarity 87.5%; Pred. No. 0.73;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAAAAA 25
|||||
Db 222 EAAAAAAAAAAAAAAAAAAAAA 245

RESULT 5

Q8BJK2 ID Q8BJK2 PRELIMINARY; PRT; 301 AA.
AC Q8BJK2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Call division cycle 2-like 5 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK083577; BAC38957.1; --
FT NON_TER 1
SQ SEQUENCE 301 AA; 32269 MW; A6CF891DBE25E09E CRC64;
Query Match 73.4%; Score 80; DB 11; Length 301;
Best Local Similarity 83.3%; Pred. No. 0.93;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAAAAA 25
|||||
Db 74 EAAAAAAAAAAAAAAAAAAAAA 97

RESULT 6

Q9EPW8 ID Q9EPW8 PRELIMINARY; PRT; 1354 AA.
AC Q9EPW8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nischarin.
GN NISCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR outbred; TISSUE=Brain;
RX MEDLINE=20571837; PubMed=11121431;
RA Alahari S.K., Lee J.W., Juliano R.L.;
RT "Nischarin, a Novel Protein That Interacts with the Integrin α 5
RT Subunit and Inhibits Cell Migration.";
RL J. Cell Biol. 151:1141-1154(2000).
DR EMBL; AF315344; AAG42100.1; --
DR MGD; MGI:128323; Nisch.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007092; LRR_SDS22.
DR Pfam; PF00560; LRR_5.
DR PRINTS; PR00019; LEURICRPT.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PROSITE; PS00504; LRR_SDS22; 1.
SQ SEQUENCE 1354 AA; 148060 MW; 01BD676FDC19247 CRC64;

Query Match 73.4%; Score 80; DB 11; Length 1354;
Best Local Similarity 83.3%; Pred. No. 3.4;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAAAAA 25
|||||
Db 837 EAPAAEAPAAEAPAAEAPAA 860

RESULT 7

Q91WW0 ID Q91WW0 PRELIMINARY; PRT; 110 AA.
AC Q91WW0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 10.7 kDa protein.
GN AI591529.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RA Strausberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC012681; AHH12681.1; -.
 DR MGD; MGI:2139008; A1591529.
 DR InterPro; IPR00104; Antifreeze 1.
 DR InterPro; IPR001859; Ribosomal_P2.
 DR PRINTS; PR00308; ANIFREEZE1.
 DR PRINTS; PR00456; RIBOSOMALP2.
 KW Hypothetical protein.
 SQ SEQUENCE 110 AA; 10662 MW; 0581D2F635F87EAB CRC64;

Query Match 72.5%; Score 79; DB 11; Length 110;
 Best Local Similarity 83.3%; Pred. No. 0.49;
 Matches 20; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 EAAAAAEEEEAAAAEEEEAAAA 25
 :|||||:|||||:|||||:|||||
 Db 23 DAAAAAEEEEAAAAEEEEAAAA 46

RESULT 8

Q81ZU1 PRELIMINARY; PRT; 332 AA.
 ID Q81ZU1
 AC Q81ZU1
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Family with sequence similarity 9, member A.
 GN FAM9A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22202142; PubMed=12213195;
 RA Martinez-Garay I., Jablonka S., Sutajova M., Steuarnagel P., Gal A.,
 RA Kutsche K.;
 RT "A New Gene Family (FAM9) of Low-Copy Repeats in Xp22.3 Expressed
 RT Exclusively in Testis: Implications for Recombinations in This
 RT Region.";
 RL Genomics 80:259-267(2002).
 DR EMBL: AF494343; AAN07162.1; -.
 SQ SEQUENCE 332 AA; 37339 MW; 92F22EC36038229C CRC64;

Query Match 72.5%; Score 79; DB 4; Length 332;
 Best Local Similarity 83.3%; Pred. No. 1.3;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 EAAAAAEEEEAAAAEEEEAAAA 25
 ||:|||||:|||||:|||||
 Db 190 EAEAEAAAAAEEEEAAAAEEEE 213

RESULT 9

Q39598 PRELIMINARY; PRT; 265 AA.
 ID Q39598
 AC Q39598
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Cgcr-4 product (Fragment).
 GN CGCR-4.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=92119224; PubMed=1731966;
 RA Wakarchuk W.W., Muller F.W., Beck C.F.;
 RT "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex
 RT arrangements of directly repeated sequence motifs.";
 RL Plant Mol. Biol. 18:143-146(1992).
 DR EMBL: X17208; CAA35080.1; -.
 FT NON TER 1
 SQ SEQUENCE 265 AA; 26216 MW; B35318B7377CF782 CRC64;

Query Match 70.6%; Score 77; DB 10; Length 265;
 Best Local Similarity 82.6%; Pred. No. 1.6;
 Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AAAAAAEEEEAAAAEEEEAAAA 25
 |||||:|||||:|||||:|||||
 Db 154 AAAAAAEEEEAAAAEEEEAAAA 176

RESULT 10

Q90XG2 PRELIMINARY; PRT; 665 AA.
 ID Q90XG2
 AC Q90XG2
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Choroideremia protein.
 GN CHM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RN STRAIN=BALB/c;
 RA van den Hurk J.A., Huber I., van de Pol T.J., Cremers F.P.;
 RT "Cloning and sequencing of the mouse choroideremia gene.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF218084; AAF25478.1; -.
 DR MGD; MGI:892979; Chm.
 DR InterPro; IPR002005; Rab_GDI_REP.
 DR Pfam; PF00996; GDI_1.
 DR PRINTS; PR00891; RABGDIREP.
 SQ SEQUENCE 665 AA; 73976 MW; FF71A74AD3FDE0A CRC64;

Query Match 69.7%; Score 76; DB 11; Length 665;
 Best Local Similarity 79.2%; Pred. No. 4.4;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 EAAAAAEEEEAAAAEEEEAAAA 25
 |||||:|||||:|||||:|||||
 Db 132 EAAEAAEAAEAAEAAEAAEAA 155

RESULT 11

Q9H4A0 PRELIMINARY; PRT; 1452 AA.
 ID Q9H4A0
 AC Q9H4A0
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CDC2L5 protein kinase.
 GN CDC2L5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,
 RA Picard A., Callebaut I., Perre E., Genevriere A.M.;
 RT "A new subfamily of high molecular mass CDC2-related kinases with

DT	01-MAY-2000 (TrEMBLrel. 13, Created)
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE	CG13012 protein.
GN	CG13012.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
SEQUENCE FROM N.A.	
RP	STRAIN=BKELYE;
RC	MEDLINE=20196006; PubMed=10731132;
RX	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA	Branan K.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA	Ballew R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J.J., Brokstein P., Brottier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA	Foster C., Gabrieli A.S., Garg N.S., Gelbart W.M., Glaesser K.,
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA	Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J.R., Puri V., Reese M.G.,
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT	"The genome sequence of Drosophila melanogaster.";
RL	Science 287:2185-2195(2000).
DR	EMB AE003503; AAF48641.1; -
DR	Flybase; FBgn030769; CG13012.
SQ	SEQUENCE 221 AA; 22987 MW; AIB95919B167C5E2 CRC64;
Query Match	68.8%; Score 75; DB 5; Length 221;
Best Local Similarity	81.8%; Pred. No. 2.1;
Matches 18; Conservative	0; Mismatches 4; Indels 0; Gaps
QY	4 AAAAAAAAAAAAAAEEAAAAAA 25
DB	6 AAAAAAAAAEAATVAEAAA 27
RESULT 14	
ID Q9W2J2	PRELIMINARY; PRT; 997 AA.
AC Q9W2J2	
DT 01-MAY-2000 (TrEMBLrel. 13, Created)	
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)	
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE CG19375 protein.	

GN CG18375.
 OC Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkelley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J., Blazey R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Bereman B.P., Bhaktar D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Fogler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J.J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorset V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003453; AAP46699.2; -;
 DR HSSP; Q13625; 1YCS
 DR FlyBase; FBGN0034606; CG18375.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR00104; Antifreeze_1.
 DR Pfam; PF00023; ank; 2.
 DR PRINTS; PR01415; ANKYRIN.
 DR PRINTS; PR00308; ANTIFREEZE1.
 DR PRODOM; PD000066; SH3; 1.
 DR SMART; SM00248; ANK; 2.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PSS0088; ANK_REPEAT; 2.
 DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PSS0002; SH3; 1.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 997 AA; 107821 MW; E712D400C2C4FD3D CRC64;
 Query Match 66.1%; Score 72; DB 5; Length 997;
 Best Local Similarity 79.2%; Pred. No. 15;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 EAAAAAEEAAAAAEEAAAAA 25
 DB 432 EAAAAAEEAAAAAEEAAAAA 455
 RESULT 15
 Q39597 PRELIMINARY; PRT; 206 AA.
 ID Q39597
 AC Q39597
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Cgcr-1 product (Fragment).
 GN Cgcr-1.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cw15 mt-;
 RX MEDLINE=92119224; PubMed=1731966;
 RA Wakarchuk W.W., Muller F.W., Beck C.F.;
 RT "Two GC-rich DNA elements of *Chlamydomonas reinhardtii* with complex
 RT arrangements of directly repeated sequence motifs.";
 RL Plant Mol. Biol. 18:143-146(1992).
 DR EMBL; X17207; CAA35079.1; -;
 FT NON_TER 1 1
 FT NON_TER 206 206
 SQ SEQUENCE 206 AA; 19869 MW; ED3FF120EF8EFAE1 CRC64;
 Query Match 65.1%; Score 71; DB 10; Length 206;
 Best Local Similarity 75.0%; Pred. No. 4.7;
 Matches 18; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 QY 2 EAAAAAEEAAAAAEEAAAAA 25
 DB 49 EAAAAAEEAAAAAEEAAAAA 72

Search completed: January 30, 2004, 00:24:39
Job time : 27.8216 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 00:13:12 ; Search time 10.2535 Seconds
(without alignments)
225.098 Million cell updates/sec

Title: US-09-461-684C-3
Perfect score: 143
Sequence: 1 GLFGAIAAGFIENGWEGMIDGWYG 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	134	93.7	550	1 HMIVS2	hemagglutinin prec
2	134	93.7	550	1 HMIVS3	hemagglutinin prec
3	134	93.7	550	1 HMIV77	hemagglutinin prec
4	134	93.7	550	1 HMIV80	hemagglutinin prec
5	134	93.7	550	1 HMIV33	hemagglutinin prec
6	134	93.7	550	1 HMIV89	hemagglutinin prec
7	134	93.7	550	1 HMIV21	hemagglutinin prec
8	134	93.7	550	1 HMIV98	hemagglutinin prec
9	134	93.7	550	1 HMIV15	hemagglutinin prec
10	134	93.7	550	2 JQ1153	hemagglutinin prec
11	134	93.7	550	2 JQ1154	hemagglutinin prec
12	134	93.7	550	2 JQ1155	hemagglutinin prec
13	134	93.7	566	1 HMIVH	hemagglutinin prec
14	134	93.7	566	1 HMIVHA	hemagglutinin prec
15	134	93.7	566	1 HMIVHM	hemagglutinin prec
16	134	93.7	566	1 HMIVDU	hemagglutinin prec
17	133	93.0	561	1 HMIV49	hemagglutinin prec
18	133	93.0	561	1 HMIV84	hemagglutinin prec
19	132	92.3	565	1 HMIVE1	hemagglutinin prec
20	132	92.3	565	1 HMIVE3	hemagglutinin prec
21	132	92.3	566	1 HMIV6	hemagglutinin prec
22	132	92.3	567	1 HMIVV	hemagglutinin prec
23	131	91.6	362	2 S38637	hemagglutinin - in
24	131	91.6	550	1 HMIV86	hemagglutinin prec
25	131	91.6	560	1 HMIVT7	hemagglutinin prec
26	131	91.6	565	1 HMIVE4	hemagglutinin prec
27	131	91.6	565	1 HMIVE5	hemagglutinin prec
28	131	91.6	565	1 HMIVE4	hemagglutinin prec
29	131	91.6	565	1 HMIVE6	hemagglutinin prec

30	131	91.6	565	1 HMIVE7	hemagglutinin prec
31	131	91.6	565	1 HMIVE8	hemagglutinin prec
32	131	91.6	565	1 HMIVE9	hemagglutinin prec
33	131	91.6	565	1 HMIVE1	hemagglutinin prec
34	131	91.6	565	1 HMIVE	hemagglutinin prec
35	131	91.6	565	2 S33703	hemagglutinin - in
36	131	91.6	570	1 A45591	hemagglutinin prec
37	131	91.6	570	2 S22013	hemagglutinin prec
38	131	91.6	570	2 S22014	hemagglutinin prec
39	131	91.6	570	2 S22015	hemagglutinin prec
40	131	91.6	570	2 S22016	hemagglutinin prec
41	131	91.6	570	2 S22017	hemagglutinin prec
42	131	91.6	570	2 S22018	hemagglutinin prec
43	131	91.6	570	2 S22020	hemagglutinin prec
44	131	91.6	570	2 S22021	hemagglutinin prec
45	131	91.6	570	2 S22029	hemagglutinin prec

ALIGNMENTS

RESULT 1

HMIVS2
hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)
C:Species: influenza A virus
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: A29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Virolgy 162, 160-166, 1988
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A:Reference number: A94370; MUID:88101364; PMID:3336940
A:Accession: A29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M19056; NID:g324208
A:Note: the sequence in GenBank entry FLAHAPA, release 106, (PID:g324209) differs from t1
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domains: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7% ; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10; Mismatches 0; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 2

HMIVS3
hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)
C:Species: influenza A virus
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: B29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Virolgy 162, 160-166, 1988
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A:Reference number: A94370; MUID:88101364; PMID:3336940
A:Accession: B29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M19057; NID:g324210
A:Note: the sequence in GenBank entry FLAHAPB, release 106, (PID:g324211) differs from t1
C:Genetics:
A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HAI #status predicted <HA1>
F:300-536/Product: hemagglutinin HAI #status predicted <HA2>
F:520-536/Product: hemagglutinin HAI #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 3
HMIV77
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)
N:Contains: hemagglutinin HAI; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A27813
R:Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; PMID:87265458; PMID:2440178
A:Accession: A27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16737; NID:G324081; PIDN:AAA43143.1; PID:G324082
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HAI #status predicted <HA1>
F:300-536/Product: hemagglutinin HAI #status predicted <HA2>
F:520-536/Product: hemagglutinin HAI #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 4
HMIV80
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)
N:Contains: hemagglutinin HAI; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C:Accession: B27813
R:Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; PMID:87265458; PMID:2440178
A:Accession: B27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16738; NID:G324083
A:Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HAI #status predicted <HA1>
F:300-536/Product: hemagglutinin HAI #status predicted <HA2>
F:520-536/Product: hemagglutinin HAI #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

F:1-328/Product: hemagglutinin HAI #status predicted <HA1>
F:300-536/Product: hemagglutinin HAI #status predicted <HA2>
F:520-536/Product: hemagglutinin HAI #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 5
HMIV33
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)
N:Contains: hemagglutinin HAI; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: C27813
R:Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; PMID:87265458; PMID:2440178
A:Accession: C27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16739; NID:G324085; PIDN:AAA43145.1; PID:G324086
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HAI #status predicted <HA1>
F:300-536/Product: hemagglutinin HAI #status predicted <HA2>
F:520-536/Product: hemagglutinin HAI #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 6
HMIV89
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)
N:Contains: hemagglutinin HAI; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: D27813
R:Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; PMID:87265458; PMID:2440178
A:Accession: D27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16740; NID:G324087; PIDN:AAA43146.1; PID:G324088
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HAI #status predicted <HA1>
F:300-536/Product: hemagglutinin HAI #status predicted <HA2>
F:520-536/Product: hemagglutinin HAI #status predicted <TM1>

F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAAGFIENGWEGMIDGWYG 352

RESULT 7

HMIV21 hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)

N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: E27813

R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: E27813

A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16741; NID:G324089.

C;Genetics:

A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin

C;Keywords: glycoprotein; hemagglutinin; homotrimer

F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F;520-536/Domains: transmembrane #status predicted <TM1>

F;7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAAGFIENGWEGMIDGWYG 352

RESULT 8

HMIV98

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)

N;Contains: hemagglutinin HA1; hemagglutinin HA2

C;Species: influenza A virus

C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998

C;Accession: F27813

R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A;Reference number: A94363; MUID:87265458; PMID:2440178

A;Accession: F27813

A;Molecule type: genomic RNA

A;Residues: 1-550 <KID>

A;Cross-references: GB:M16742; NID:G324091

C;Genetics:

A;Map position: segment 4

C;Superfamily: influenza virus hemagglutinin

C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F;520-536/Domains: transmembrane #status predicted <TM1>

F;7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAAGFIENGWEGMIDGWYG 352

RESULT 9

HMIV15

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)

N;Contains: hemagglutinin HA1; hemagglutinin HA2

C;Species: influenza A virus

C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999

C;Accession: G27813

R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A;Reference number: A94363; MUID:87265458; PMID:2440178

A;Accession: G27813

A;Molecule type: genomic RNA

A;Residues: 1-550 <KID>

A;Cross-references: GB:M16743; NID:G324093; PIDN:AAA43149.1; PID:G324094

C;Genetics:

A;Map position: segment 4

C;Superfamily: influenza virus hemagglutinin

C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F;520-536/Domains: transmembrane #status predicted <TM1>

F;7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAAGFIENGWEGMIDGWYG 352

RESULT 10

JQ1153

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)

N;Contains: hemagglutinin HA1; hemagglutinin HA2

C;Species: influenza A virus

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000

C;Accession: JQ1153

R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.

J. Gen. Virol. 72, 2007-2010, 1991

A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3

A;Reference number: JQ1153; MUID:91341491; PMID:1875195

A;Accession: JQ1153

A;Molecule type: genomic RNA

A;Residues: 1-550 <YAS>

A;Cross-references: GB:D00929; NID:G221279; PIDN:BAA00769.1; PID:G221280

A;Note: the authors translated the codon GGG for residue 218 as Glu

A;Note: residues 528-532 are not shown in this publication

C;Superfamily: influenza virus hemagglutinin

C;Keywords: glycoprotein; homotrimer

F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
|||||

Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 11

QJ1154
hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)
N/Contains: hemagglutinin HA1; hemagglutinin HA2
C/Species: influenza A virus
C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C/Accession: JQ1154
R/Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A/Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3N2 virus
A/Reference number: JQ1153; MUID:91341491; PMID:1875195
A/Accession: JQ1154
A/Molecule type: genomic RNA
A/Residues: 1-550 <YAS>
A/Cross-references: GB:D00930; NID:G221273; PIDN:BAA00770.1; PID:G221274
A/Note: the authors translated the codon GGG for residue 218 as Glu
C/Keywords: influenza A virus
C/Superfamily: influenza virus hemagglutinin
C/Keywords: glycoprotein; homotrimer
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 12

QJ1155
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)
N/Contains: hemagglutinin HA1; hemagglutinin HA2
C/Species: influenza A virus
C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C/Accession: JQ1155
R/Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A/Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3N2 virus
A/Reference number: JQ1153; MUID:91341491; PMID:1875195
A/Accession: JQ1155
A/Molecule type: genomic RNA
A/Residues: 1-550 <YAS>
A/Cross-references: GB:D00931; NID:G221277; PIDN:BAA00771.1; PID:G221278
A/Note: the authors translated the codon GGG for residue 218 as Glu, GCC for residue 538
A/Note: residues 528-532 are not shown in this publication
C/Superfamily: influenza virus hemagglutinin
C/Keywords: glycoprotein; homotrimer
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 13

HM1VHA
hemagglutinin precursor - influenza A virus
C/Species: influenza A virus
C/Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 22-Oct-1999
C/Accession: A93705; A93233; A04051; A93231; A94441

R/Both, G.W.; Sleight, M.J.
Nucleic Acids Res. 8, 2561-2575, 1980
A/Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza A virus
A/Reference number: A93705; MUID:81053698; PMID:6253883
A/Accession: A93705
A/Molecule type: genomic RNA
A/Residues: 1-566 <BOT>
A/Cross-references: GB:V01103
A/Experimental source: strain A/NT/60/68/29C
A/Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/NT/60/68/29C
R/Dopheide, T.A.; Ward, C.W.
FEBS Lett. 110, 181-183, 1980
A/Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.
A/Reference number: A91276; MUID:80179105; PMID:6768586
A/Contents: annotation; disulfide bonds
R/Gething, M.J.; Bye, J.; Skehel, J.; Waterfield, M.
Nature 287, 301-306, 1980
A/Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from influenza A virus
A/Reference number: A93233; MUID:81030852; PMID:7421990
A/Accession: A93233
A/Molecule type: genomic RNA
A/Residues: 1-24,'S',26,'D',28-159,'G',161-197,'I',199-241,'L',243-249 <GET>
A/Experimental source: strain X-31[H3]
C/Superfamily: influenza virus hemagglutinin
C/Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:536-552/Domain: transmembrane #status predicted <TM1>
F:30-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status experimental
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 346 GLFGAIAAGFIENGWEGMIDGWYG 368

RESULT 14

HM1VHA
hemagglutinin precursor - influenza A virus (strain A/Aichi/2/68)
N/Contains: hemagglutinin HA1; hemagglutinin HA2
C/Species: influenza A virus
C/Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C/Accession: A93231; A04051
R/Verhoeven, M.; Fang, R.; Min Jou, W.; Devos, R.; Huylebroeck, D.; Saman, E.; Fiers, W.
Nature 286, 771-776, 1980
A/Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/England/1968 and A/England/1969
A/Reference number: A93231; MUID:80254693; PMID:7402351
A/Accession: A93231
A/Molecule type: genomic RNA
A/Residues: 1-566 <VER>
A/Cross-references: GB:J02090; NID:G324131; PIDN:AAA43178.1; PID:G324132
C/Superfamily: influenza virus hemagglutinin
C/Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 346 GLFGAIAAGFIENGWEGMIDGWYG 368

RESULT 15

HM1VEM
hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: Influenza A virus
C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 31-Mar-2000
C:Accession: A94441; A04051
R:Sleigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.
in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, B
A:Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of c
A:Reference number: A94441
A:Accession: A94441
A:Molecule type: genomic RNA
A:Residues: 1-566 <SLE>
C:Superfamily: influenza virus hemagglutinin
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
Db 346 GLFGAIAGFIENGWEGMIDGWYG 368

Search completed: January 30, 2004, 00:26:21
Job time : 10.2535 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:27 ; Search time 5.40845 Seconds
(without alignments)
208.681 Million cell updates/sec

Title: US-09-461-684C-3

Perfect score: 143

Sequence: 1 CGLFGAIAPIENGWEGMIDGWYG 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	93.7	550	1 HEMA_IADH1	P12582 influenza a
2	134	93.7	550	1 HEMA_IADH2	P12583 influenza a
3	134	93.7	550	1 HEMA_IADH3	P12584 influenza a
4	134	93.7	550	1 HEMA_IADH4	P12585 influenza a
5	134	93.7	550	1 HEMA_IADH5	P12586 influenza a
6	134	93.7	550	1 HEMA_IADH6	P12587 influenza a
7	134	93.7	550	1 HEMA_IADH7	P12588 influenza a
8	134	93.7	550	1 HEMA_IADH8	P43257 influenza a
9	134	93.7	550	1 HEMA_IADH9	P43258 influenza a
10	134	93.7	550	1 HEMA_IADH10	P43260 influenza a
11	134	93.7	550	1 HEMA_IADH11	P11133 influenza a
12	134	93.7	550	1 HEMA_IADH12	P11134 influenza a
13	134	93.7	566	1 HEMA_IADH13	P03437 influenza a
14	134	93.7	566	1 HEMA_IADH14	P261134 influenza a
15	134	93.7	566	1 HEMA_IADH15	P03442 influenza a
16	134	93.7	566	1 HEMA_IADH16	P261138 influenza a
17	134	93.7	566	1 HEMA_IADH17	P03449 influenza a
18	134	93.7	566	1 HEMA_IADH18	P03439 influenza a
19	134	93.7	566	1 HEMA_IADH19	P03436 influenza a
20	133	93.0	561	1 HEMA_IADH20	P12581 influenza a
21	133	93.0	561	1 HEMA_IADH21	P12439 influenza a
22	132	92.3	565	1 HEMA_IADH22	P17000 influenza a
23	132	92.3	565	1 HEMA_IADH23	P17002 influenza a
24	132	92.3	566	1 HEMA_IADH24	P03440 influenza a
25	132	92.3	566	1 HEMA_IADH25	P261139 influenza a
26	132	92.3	567	1 HEMA_IADH26	P03435 influenza a
27	131	91.6	550	1 HEMA_IADH27	P12589 influenza a
28	131	91.6	560	1 HEMA_IADH28	P03458 influenza a
29	131	91.6	565	1 HEMA_IADH29	P16994 influenza a
30	131	91.6	565	1 HEMA_IADH30	P16995 influenza a
31	131	91.6	565	1 HEMA_IADH31	P19699 influenza a
32	131	91.6	565	1 HEMA_IADH32	P16996 influenza a
33	131	91.6	565	1 HEMA_IADH33	P15658 influenza a

ALIGNMENTS

RESULT 1

HEMA_IADH1 STANDARD; PRT; 550 AA.
 ID HEMA_IADH1
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DB Hemagglutinin precursor (Contains: Hemagglutinin HA1 chain;
 DB Hemagglutinin HA2 chain) (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/5/77).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11357;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawaoaka Y., Naeye C.W., Webster R.G.;
 RT "Antigenic and Genetic conservation of H3 influenza virus in wild
 ducks.";
 RL Virology 159:109-119(1987).
 CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; M16737; AAA43143.1; --
 CC HSSP; P03437; 3HMG.
 DR InterPro; IPR001384; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON_TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61705 MW; 7E7ACFE716FC969A CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 1.2e-10; Mismatches 0; Indels 0; Gaps 0;
 Matches 23; Conservative 0;

```

QY      2 GLFGAIAAGFIENGWEGMIDGWYG 24
DB      330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 2
HEMA_IADH2
ID      HEMA_IADH2      STANDARD;      PRT;      550 AA.
AC      P12583; Q84011;
DT      01-OCT-1989 (Rel. 12, Created)
DT      01-OCT-1989 (Rel. 14, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE      Hemagglutinin HA2 chain] (Fragment).
GN      HA.
OS      Influenza A virus (strain A/Duck/Hokkaido/8/80).
OC      Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC      Influenza A viruses; Influenzavirus A.
OX      NCBI_TaxID=11358;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=87265458; PubMed=2440178;
RX      Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT      "Antigenic and genetic conservation of H3 influenza virus in wild
RT      ducks.";
RL      Virology 159:109-119(1987).
CC      -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC      CELL RECEPTORS AND FOR INITIATING INFECTION.
CC      -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC      (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC      -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M16738; AAA43144.1; -.
CC      PIR; B27813; HMV80.
CC      HSSP; P03437; 2VIU.
CC      InterPro; IPR001364; Hemagglutn.
CC      Pfam; PF00509; Hemagglutinin; 1.
CC      PRINTS; PR00329; HEMAGGLUTN12.
CC      ProDom; PD000225; Hemagglutn; 1.
CC      Envelope protein; Hemagglutinin; Glycoprotein.
CC      NON_TER      1
CC      CHAIN      1 328      HEMAGGLUTININ HA1 CHAIN.
CC      FT      CHAIN      330 550      HEMAGGLUTININ HA2 CHAIN.
CC      FT      CARBOHYD      8 8      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      22 22      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      38 38      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      165 165      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      285 285      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      483 483      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CONFLICT      137 137      K -> N (IN PIR DATA BANK).
CC      SQ      SEQUENCE      550 AA; 61659 MW; A107023ACCC9CC353 CRC64;

Query Match      93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GLFGAIAAGFIENGWEGMIDGWYG 24
DB      330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 3
HEMA_IADH3
ID      HEMA_IADH3      STANDARD;      PRT;      550 AA.
AC      P12585; Q84013; Q84014;
DT      01-OCT-1989 (Rel. 12, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE      Hemagglutinin HA2 chain] (Fragment).
GN      HA.
OS      Influenza A virus (strain A/Duck/Hokkaido/7/82).
OC      Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC      Influenza A viruses; Influenzavirus A.
OX      NCBI_TaxID=11360;

```

```

AC      P12584; Q84012; Q89793;
DT      01-OCT-1989 (Rel. 12, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE      Hemagglutinin HA2 chain] (Fragment).
GN      HA.
OS      Influenza A virus (strain A/Duck/Hokkaido/33/80).
OC      Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC      Influenza A viruses; Influenzavirus A.
OX      NCBI_TaxID=11359;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=87265458; PubMed=2440178;
RX      Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT      "Antigenic and genetic conservation of H3 influenza virus in wild
RT      ducks.";
RL      Virology 159:109-119(1987).
CC      -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC      CELL RECEPTORS AND FOR INITIATING INFECTION.
CC      -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC      (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC      -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M16739; AAA43145.1; -.
CC      HSSP; P03437; 2VIU.
CC      InterPro; IPR001364; Hemagglutn.
CC      Pfam; PF00509; Hemagglutinin; 1.
CC      PRINTS; PR00329; HEMAGGLUTN12.
CC      ProDom; PD000225; Hemagglutn; 1.
CC      Envelope protein; Hemagglutinin; Glycoprotein.
CC      NON_TER      1
CC      CHAIN      1 328      HEMAGGLUTININ HA1 CHAIN.
CC      FT      CHAIN      330 550      HEMAGGLUTININ HA2 CHAIN.
CC      FT      CARBOHYD      8 8      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      22 22      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      38 38      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      165 165      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      285 285      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      483 483      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      SQ      SEQUENCE      550 AA; 61577 MW; 6C30BF67CFDCB7DE CRC64;

Query Match      93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GLFGAIAAGFIENGWEGMIDGWYG 24
DB      330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 4
HEMA_IADH4
ID      HEMA_IADH4      STANDARD;      PRT;      550 AA.
AC      P12585; Q84013; Q84014;
DT      01-OCT-1989 (Rel. 12, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE      Hemagglutinin HA2 chain] (Fragment).
GN      HA.
OS      Influenza A virus (strain A/Duck/Hokkaido/7/82).
OC      Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC      Influenza A viruses; Influenzavirus A.
OX      NCBI_TaxID=11360;

```

RN SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawakita Y., Naevae C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RL ducks";
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M16740; AAA43146.1; -
 DR HSP; P03437; 2VIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 178 179 YV -> VI (IN PIR DATA BANK).
 FT CONFLICT 388 388 K -> T (IN PIR DATA BANK).
 SQ SEQUENCE 550 AA; 61664 MW; A16B2CF8CBBD900 CRC64;
 Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLFGATAGFIENGWEGMIDGWY 24
 Db 330 GLFGATAGFIENGWEGMIDGWY 352
 RESULT 5
 HEMA IADH5 STANDARD; PRT; 550 AA.
 AC P12586; Q84015; Q84016;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11361;
 RN SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawakita Y., Naevae C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RL ducks";
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M16740; AAA43146.1; -
 DR HSP; P03437; 2VIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 178 179 YV -> VI (IN PIR DATA BANK).
 FT CONFLICT 388 388 K -> T (IN PIR DATA BANK).
 SQ SEQUENCE 550 AA; 61664 MW; A16B2CF8CBBD900 CRC64;
 Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLFGATAGFIENGWEGMIDGWY 24
 Db 330 GLFGATAGFIENGWEGMIDGWY 352

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M16741; AAA43147.1; -
 DR PIR; E27813; HMIV21.
 DR HSP; P03437; 2VIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 7 7 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 178 179 YV -> VI (IN PIR DATA BANK).
 FT CONFLICT 388 388 K -> T (IN PIR DATA BANK).
 SQ SEQUENCE 550 AA; 61856 MW; 48401C867A15BF8C CRC64;
 Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLFGATAGFIENGWEGMIDGWY 24
 Db 330 GLFGATAGFIENGWEGMIDGWY 352
 RESULT 6
 HEMA IADH6 STANDARD; PRT; 550 AA.
 AC P12587; Q84017;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11362;
 RN SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawakita Y., Naevae C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RL ducks";
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC
CC
CC
DR EMBL; M16742; AAA43148.1; -
DR PIR; F27813; HMIV98.
DR HSSP; P03437; LHGJ.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 8
SQ SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CPE CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 7

HEMA_IADH7
ID HEMA_IADH7 STANDARD; PRT; 550 AA.
AC P12588; Q84018; Q89470;
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/10/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11363;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawachi Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119 (1987).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC
CC EMBL; M16743; AAA43149.1; -
CC HSSP; P03437; 3HMG.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61761 MW; 65F81793281D53EB CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 8

HEMA_IADHK
ID HEMA_IADHK STANDARD; PRT; 550 AA.
AC P43257;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hong Kong/7/75).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11364;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/Hong Kong/68 (H3N2) strain emerged.";
RL J. Gen. Virol. 72:2007-2010 (1991).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; D00929; BAA00769.1; -
CC HSSP; P03437; 2VIU.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61549 MW; 864639B829FE1BA9 CRC64;


```
Query Match          93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
   |||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 9
HEMA IADHL
ID _HEMA_IADHL STANDARD; PRT; 550 AA.
AC P43258;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hong Kong/64/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=45412;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/Hong Kong/68 (H3N2) strain emerged.";
RL J. Gen. Virol. 72:2007-2010(1991).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00931; BAA00771.1; --
CC HSP; P03437; 2VIU.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein.
CC NON_TER 1 328 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 550 AA; 61718 MW; A351C56789E4BE9A CRC64;

Query Match          93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
   |||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 11
HEMA IAZH2
ID _HEMA_IAZH2 STANDARD; PRT; 550 AA.
AC P11133; Q84019; Q84020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE chain] (Fragment).
GN HA.
```

```
RESULT 10
HEMA IAGHK
ID _HEMA_IAGHK STANDARD; PRT; 550 AA.
AC P43260;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Goose/Hong Kong/10/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=45414;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/Hong Kong/68 (H3N2) strain emerged.";
RL J. Gen. Virol. 72:2007-2010(1991).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00930; BAA00770.1; --
CC HSP; P03437; 2VIU.
CC InterPro; IPR001364; Hemagglutn.
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC ProDom; PD000225; Hemagglutn; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein.
CC NON_TER 1 328 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
CC CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 550 AA; 61676 MW; 9A1E094DA28BACD2 CRC64;

Query Match          93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
   |||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 11
HEMA IAZH2
ID _HEMA_IAZH2 STANDARD; PRT; 550 AA.
AC P11133; Q84019; Q84020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE chain] (Fragment).
GN HA.
```

OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
 OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11497;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88101364; PubMed=3336940;
 RA Kida H., Shortridge K.F., Webster R.G.;
 RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
 in China.";
 RL Virology 162:160-166(1988).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL-RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M19057; AAA43212.1; -;
 CC HSSP; P03437; 2VIU.
 CC InterPro; IPR001364; Hemagglutn.
 CC Pfam; PF00509; Hemagglutinin; 1.
 CC PRINTS; PR00329; Hemagglutn12.
 CC ProDom; PD000225; Hemagglutn; 1.
 CC Hemagglutinin; Envelope protein; Glycoprotein.
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61437 MW; 1F2A7E758C531CE8 CRC64;
 Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLFGAAGTENGWEGMIDGWYG 24
 DB 330 GLFGAAGTENGWEGMIDGWYG 352
 RESULT 12
 HEMA_I2H3
 ID HEMA_I2H3 STANDARD; PRT; 550 AA.
 AC P11134; Q84025; Q84026;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
 DE chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
 OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11498;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88101364; PubMed=3336940;
 RA Kida H., Shortridge K.F., Webster R.G.;
 RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
 in China.";
 RL Virology 162:160-166(1988).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL-RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M19056; AAA43211.1; ALT_TERM.
 CC HSSP; P03437; 2VIU.
 CC InterPro; IPR001364; Hemagglutn.
 CC Pfam; PF00509; Hemagglutinin; 1.
 CC PRINTS; PR00329; Hemagglutn12.
 CC ProDom; PD000225; Hemagglutn; 1.
 CC Hemagglutinin; Envelope protein; Glycoprotein.
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61580 MW; 991F6D8BC02F24F2 CRC64;
 Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLFGAAGTENGWEGMIDGWYG 24
 DB 330 GLFGAAGTENGWEGMIDGWYG 352
 RESULT 13
 HEMA_I2A1C
 ID HEMA_I2A1C STANDARD; PRT; 566 AA.
 AC P03437;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/Aichi/2/68).
 OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=150147;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80254693; PubMed=7402351;
 RA Verhoeven M., Fang R., Min Jou W., Devos R., Huylebroeck D.,
 RA Saman E., Fiers W.;
 RT "Antigenic drift between the haemagglutinin of the Hong Kong
 RT influenza strains A/Aichi/2/68 and A/Victoria/3/75.";
 RL Nature 286:771-776(1980).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RX MEDLINE=81123029; PubMed=7464906;
 RA Wilson I.A., Skehel J.J., Wiley D.C.;
 RT "Structure of the haemagglutinin membrane glycoprotein of influenza
 RT virus at 3-A resolution.";
 RL Nature 289:366-373(1981).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY.
 RX MEDLINE=88232903; PubMed=3374584;

RA Weis W.I., Brown J.H., Cusack S.C., Paulson J.C., Skehel J.J.,
 RA Wiley D.C.;
 RT "Structure of the influenza virus haemagglutinin complexed with its
 RT receptor, sialic acid.";
 RL Nature 333:426-431(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY OF A MUTANT WITH GLY-457.
 RX MEDLINE=90107940; PubMed=2295311;
 RA Weis W.I., Cusack S.C., Brown J.H., Daniels R.S., Skehel J.J.,
 RA Wiley D.C.;
 RT "The structure of a membrane fusion mutant of the influenza virus
 RT haemagglutinin.";
 RL EMBO J. 9:17-24(1990).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY.
 RX MEDLINE=90230310; PubMed=2329580;
 RA Weis W.I., Bruenger A.T., Skehel J.J., Wiley D.C.;
 RT "Refinement of the influenza virus haemagglutinin by simulated
 RT annealing.";
 RL J. Mol. Biol. 212:737-761(1990).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=94352388; PubMed=8072525;
 RA Bullough P.A., Hughson F.M., Skehel J.J., Wiley D.C.;
 RT "Structure of influenza haemagglutinin at the pH of membrane fusion.";
 RL Nature 371:37-43(1994).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (3.25 ANGSTROMS).
 RX MEDLINE=98120975; PubMed=9461077;
 RA Fleury D., Wharton S.A., Skehel J.J., Knossow M., Bizebard T.;
 RT "Antigen distortion allows influenza virus to escape neutralization.";
 RL Nat. Struct. Biol. 5:119-123(1998).
 CC -I- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -I- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -I- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: J02090; RAA43178.1; --
 CC EMBL: V01085; CAA24269.1; --
 CC PIR: A93231; HMI1VHA.
 CC DR PDB; 2HMG; 31-OCT-93.
 CC DR PDB; 3HMG; 31-OCT-93.
 CC DR PDB; 4HMG; 31-OCT-93.
 CC DR PDB; 5HMG; 31-JAN-94.
 CC DR PDB; 1HGD; 31-JAN-94.
 CC DR PDB; 1HGE; 31-JAN-94.
 CC DR PDB; 1HGF; 31-JAN-94.
 CC DR PDB; 1HGG; 31-JAN-94.
 CC DR PDB; 1HGH; 31-JAN-94.
 CC DR PDB; 1HGI; 31-JAN-94.
 CC DR PDB; 1HGJ; 31-JAN-94.
 CC DR PDB; 1HTW; 14-FEB-95.
 CC DR PDB; 2VIR; 29-APR-98.
 CC DR PDB; 2VIT; 29-APR-98.
 CC DR PDB; 2VIU; 29-APR-98.
 CC DR PDB; 1EO8; 29-NOV-00.
 CC DR PDB; 1HA0; 22-DEC-99.
 CC DR PDB; 1J8H; 13-MAR-02.
 CC DR PDB; 1KEN; 24-APR-02.
 CC DR PDB; 1QFU; 29-DEC-99.
 CC DR PDB; 1QUL; 05-JAN-00.
 CC DR InterPro; IPR001364; Hemagglutn.
 CC DR Pfam; PF00509; Hemagglutinin; 1.

DR PRINTS; PRO0329; HEMAGGLUTN12.
 DR ProDom; PD000235; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
 FT SIGNAL 1 16
 FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
 FT DISULFID 30 482 INTERCHAIN.
 FT DISULFID 68 293
 FT DISULFID 80 92
 FT DISULFID 113 155
 FT DISULFID 297 321
 FT DISULFID 489 493
 FT CARBOHYD 24 24
 FT CARBOHYD 38 38
 FT CARBOHYD 54 54
 FT CARBOHYD 97 97
 FT CARBOHYD 181 181
 FT CARBOHYD 301 301
 FT CARBOHYD 499 499
 FT STRAND 27 35
 FT STRAND 40 42
 FT STRAND 50 52
 FT STRAND 55 57
 FT STRAND 59 60
 FT STRAND 67 70
 FT STRAND 74 76
 FT TURN 78 79
 FT HELIX 82 87
 FT TURN 88 88
 FT HELIX 90 95
 FT TURN 96 97
 FT STRAND 99 99
 FT STRAND 102 105
 FT TURN 107 108
 FT STRAND 116 117
 FT TURN 119 120
 FT HELIX 121 131
 FT STRAND 133 133
 FT STRAND 136 138
 FT TURN 144 145
 FT STRAND 146 147
 FT STRAND 152 157
 FT TURN 158 159
 FT STRAND 160 162
 FT TURN 165 166
 FT STRAND 167 169
 FT STRAND 171 173
 FT TURN 174 175
 FT STRAND 176 176
 FT STRAND 180 185
 FT STRAND 192 200
 FT HELIX 204 211
 FT STRAND 217 221
 FT STRAND 226 229
 FT STRAND 239 239
 FT TURN 240 241
 FT STRAND 242 242
 FT STRAND 245 253
 FT TURN 255 256
 FT STRAND 258 265
 FT STRAND 267 270
 FT STRAND 272 275
 FT STRAND 282 285
 FT STRAND 290 294
 FT STRAND 297 299
 FT TURN 300 301
 FT STRAND 302 304
 FT STRAND 310 311
 FT STRAND 318 320
 FT STRAND 323 324
 FT STRAND 331 333
 FT STRAND 337 337
 FT TURN 347 348

FT TURN 350 350
 FT STRAND 351 351
 FT TURN 352 354
 FT STRAND 353 355
 FT STRAND 354 355
 FT STRAND 355 360
 FT TURN 356 361
 FT STRAND 357 373
 FT TURN 374 375
 FT STRAND 376 382
 FT STRAND 383 400
 FT TURN 401 401
 FT STRAND 406 407
 FT HELIX 421 471
 FT STRAND 472 474
 FT STRAND 475 477
 FT STRAND 482 485
 FT STRAND 491 498
 FT TURN 499 500
 FT HELIX 504 515
 FT TURN 518 519
 SQ SEQUENCE 566 AA; 63415 MW; E395659C23CAFECA CRC64;

Query Match 93.7%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
 |||||
 Db 346 GLFGAIAAGFIENGWEGMIDGWYG 368
 |||||

RESULT 14
 HEMA_IADA3
 ID HEMA_IADA3 STANDARD; PRT; 566 AA.
 AC P26134;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/Duck/Alberta/78/76).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11348;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92114135; PubMed=1731092;
 RA Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
 RA Webster R.G.,
 RT "Evolution of the H3 influenza virus hemagglutinin from human and
 nonhuman hosts."
 RL J. Virol. 66:1129-1138(1992).
 CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M73771; -; NOT_ANNOTATED_CDS.
 DR HSP; P03437; 2V1U.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 566 AA; 63534 MW; FE19AB6FP9415B89 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
 |||||
 Db 346 GLFGAIAAGFIENGWEGMIDGWYG 368
 |||||

RESULT 15
 HEMA_IADU3
 ID HEMA_IADU3 STANDARD; PRT; 566 AA.
 AC P03442;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/Duck/Ukraine/1/63).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11374;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8202542; PubMed=6169439;
 RA Fang R., Min Jou W., Huybrebeck D., Devos R., Fiers W.,
 RT "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin
 gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza
 hemagglutinin."
 RL Cell 25:315-323(1981).
 CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; V01087; CAA24271.1; -.
 DR PDB; 1IBN; 08-AUG-01.
 DR PDB; 1IBO; 08-AUG-01.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
 FT SIGNAL 1 16
 FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:52 ; Search time 24.7887 Seconds
(without alignments)
249.842 Million cell updates/sec

Title: US-09-461-684C-3
Perfect score: 143
Sequence: 1 CGLFGAAGFIENGWEGMIDGWYG 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL_23.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_phage.*
- 11: sp_plant.*
- 12: sp_rodent.*
- 13: sp_virus.*
- 14: sp_vertebrate.*
- 15: sp_unclassified.*
- 16: sp_rvirus.*
- 17: sp_bacteriap.*
- 18: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	93.7	384	Q8JK63	Q8JK63 influenza a
2	134	93.7	566	Q98052	Q98052 influenzavi
3	134	93.7	566	Q8U251	Q8U251 influenza a
4	134	93.7	566	Q8QUN8	Q8QUN8 influenza a
5	134	93.7	566	Q67132	Q67132 influenza a
6	134	93.7	566	Q67125	Q67125 influenzavi
7	134	93.7	566	Q8UXR3	Q8UXR3 influenza a
8	134	93.7	566	Q91MA7	Q91MA7 influenza a
9	134	93.7	566	Q9DHG0	Q9DHG0 influenza a
10	134	93.7	566	Q910M5	Q910M5 influenza a
11	134	93.7	566	Q67126	Q67126 influenzavi
12	133	93.0	301	Q9DXE3	Q9DXE3 influenza a
13	132	92.3	550	Q82499	Q82499 influenzavi
14	132	92.3	550	Q82498	Q82498 influenzavi
15	132	92.3	550	Q82753	Q82753 influenzavi
16	132	92.3	566	Q82496	Q82496 influenzavi

17	132	92.3	571	12	Q03909	Q03909 influenza a
18	131	91.6	109	12	Q67050	Q67050 influenzavi
19	131	91.6	109	12	Q67053	Q67053 influenzavi
20	131	91.6	109	12	Q67051	Q67051 influenzavi
21	131	91.6	109	12	Q67052	Q67052 influenzavi
22	131	91.6	362	12	Q9QKD3	Q9QKD3 influenza a
23	131	91.6	362	12	Q9QKD1	Q9QKD1 influenza a
24	131	91.6	362	12	Q82513	Q82513 influenzavi
25	131	91.6	362	12	Q9QKD2	Q9QKD2 influenza a
26	131	91.6	362	12	Q84174	Q84174 influenzavi
27	131	91.6	362	12	Q82517	Q82517 influenzavi
28	131	91.6	365	12	Q9DL25	Q9DL25 influenza a
29	131	91.6	367	12	Q9DL22	Q9DL22 influenza a
30	131	91.6	368	12	Q9DL29	Q9DL29 influenza a
31	131	91.6	369	12	Q9DL26	Q9DL26 influenza a
32	131	91.6	369	12	P87689	P87689 influenzavi
33	131	91.6	369	12	Q9DL06	Q9DL06 influenza a
34	131	91.6	371	12	Q9DL24	Q9DL24 influenza a
35	131	91.6	371	12	P87685	P87685 influenzavi
36	131	91.6	373	12	Q9DL20	Q9DL20 influenza a
37	131	91.6	374	12	Q9DL21	Q9DL21 influenza a
38	131	91.6	375	12	Q9DL27	Q9DL27 influenza a
39	131	91.6	375	12	Q9DL05	Q9DL05 influenza a
40	131	91.6	376	12	Q9DL30	Q9DL30 influenza a
41	131	91.6	376	12	Q9DL04	Q9DL04 influenza a
42	131	91.6	377	12	Q9E7P6	Q9E7P6 influenza a
43	131	91.6	382	12	Q9DL03	Q9DL03 influenza a
44	131	91.6	408	12	Q9E7P5	Q9E7P5 influenza a
45	131	91.6	409	12	Q9Q0L5	Q9Q0L5 influenza a

ALIGNMENTS

RESULT 1

Q8JK63 PRELIMINARY; PRT; 384 AA.
AC Q8JK63;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Hemagglutinin (Fragment).
GN H3HA.
OS Influenza A virus (A/teal/Germany/wv201r/01).
OC Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=205472;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/teal/Germany/wv01r/01;
RA Werner O., Starick E., Mueller T., Muehle R.;
RT "Characterisation of avian influenza virus isolates from wild birds from Germany."
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -|- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC EMBL; AJ506781; CAD44999.1; ...
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 384
SQ SEQUENCE 384 AA; 42076 MW; 459731795CA5CE38 CRC64;

Query Match 93.7%; Score 134; DB 12; Length 384;
Best Local Similarity 100.0%; Pred. No. 5.9e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24

```

Db      346 GLFGAIAAGFIENGWEGMIDGWYG 368
|||||
RESULT 2
Q98052      PRELIMINARY;      PRT;      566 AA.
AC Q98052;
DT 01-FEB-1997 (TRENBLrel. 02, Created)
DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Hemagglutinin precursor (Fragment).
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OC NCB1_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81053698; PubMed=62539883;
RA Both G.W., Sleight M.J.;
RT "Complete nucleotide sequence of the haemagglutinin gene from a human
RT influenza virus of the Hong Kong subtype.";
RL Nucleic Acids Res. 8:2561-2575(1980).
[2]
RP SEQUENCE OF 17-344 FROM N.A.
RX MEDLINE=81194918; PubMed=6164798;
RA Sleight M.J., Both G.W., Underwood P.A., Bender V.J.;
RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza
RT subtype: Correlation of amino acid changes with alterations in viral
RT antigenicity.";
RL J. Virol. 37:845-853(1981).
RN [3]
RP SEQUENCE OF 17-566 FROM N.A.
RX MEDLINE=8203276; PubMed=6169843;
RA Both G.W., Sleight M.J.;
RT "Conservation and variation in the hemagglutinins of Hong Kong subtype
RT influenza viruses during antigenic drift.";
RL J. Virol. 39:845-853(1981).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; J02135; AA443189.1; -.
DR HSP; F03437; IHGE.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 344 POTENTIAL.
FT CHAIN 346 566 POTENTIAL.
SQ SEQUENCE 566 AA; 63414 MW; C447FD465BE4FCF9 CRC64;

Query Match      93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
DB      346 GLFGAIAAGFIENGWEGMIDGWYG 368
|||||
RESULT 3
Q98051      PRELIMINARY;      PRT;      566 AA.
AC Q98051;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Haemagglutinin.
OS Haemagglutinin.

```

```

OS Influenza A virus (A/pet bird/Hong Kong/1559/99(H3N8)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OC NCB1_TaxID=183665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/pet bird/Hong Kong/1559/99;
RA Chin P., Shortridge K.F.;
RT "Characterisation of avian H3 influenza viruses.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ427304; CAD20336.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63403 MW; F11C91B6A0183484 CRC64;

Query Match      93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
DB      346 GLFGAIAAGFIENGWEGMIDGWYG 368
|||||
RESULT 4
Q8QLN8      PRELIMINARY;      PRT;      566 AA.
ID Q8QLN8;
AC Q8QLN8;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Haemagglutinin.
GN HA.
OS Influenza A virus (A/aquatic bird/Hong Kong/399/99(H3N8)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OC NCB1_TaxID=183664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/aquatic bird/Hong Kong/399/99;
RA Chin P., Shortridge K.F.;
RT "Characterisation of influenza viruses from wild aquatic birds.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ427297; CAD20322.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63412 MW; 68913C222C97B92E CRC64;

Query Match      93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
DB      346 GLFGAIAAGFIENGWEGMIDGWYG 368
|||||
RESULT 5

```

Q67132
ID Q67132 PRELIMINARY; PRT; 566 AA.
AC Q67132;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenza A virus (strain A/Aichi/2/68).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=150147;
RN [1] _____
RP SEQUENCE FROM N.A.
RC STRAIN=A/Aichi/2/68;
RA Min J.W., Verhoeven M., Fang R.-X., Devos R., Huylebroeck D.,
RA Fiers W.;
RT "Shift and drift in influenza viruses."
RL (In) Carlile M.J., Collins J.F., Moseley B.E. B. (eds.);
RL SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, pp.285-311,
RL Cambridge University Press, New York (1981).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; M55059; AAA43239.1; -;
DR HSSP; P03437; 1HGE.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT CHAIN 1 344 HEMAGGLUTININ.
FT CHAIN 346 566 NEURAMINIDASE.
SQ SEQUENCE 566 AA; 63441 MW; E5D1B97DF96FECA CRC64;
Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAAGFIENGWEGMIDGWYG 24
Db 346 GLFGAAGFIENGWEGMIDGWYG 368
RESULT 6
Q67125
ID Q67125 PRELIMINARY; PRT; 566 AA.
AC Q67125;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OX NCBI_TaxID=197911;
RN [1] _____
RP SEQUENCE FROM N.A.
RC STRAIN=A/Seal/MA/3911/92;
RX MEDLINE=95146951; PubMed=7844533;
RA Callan R.J., Early G., Kida H., Hinshaw V.S.;
RT "The appearance of H3 influenza viruses in seals."
RL J. Gen. Virol. 76:199-203 (1995).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; L31949; AAA4229.1; -;
DR HSSP; P03437; 2VIU.

DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63456 MW; AE556302A9EEB99F CRC64;
Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAAGFIENGWEGMIDGWYG 24
Db 346 GLFGAAGFIENGWEGMIDGWYG 368
RESULT 7
Q8UXR3
ID Q8UXR3 PRELIMINARY; PRT; 566 AA.
AC Q8UXR3;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenza A virus (A/swine/Potsdam/35/82 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=183769;
RN [1] _____
RP SEQUENCE FROM N.A.
RC STRAIN=A/swine/Potsdam/35/82;
RA Grotzinger I., Sues J., Grotzinger C.;
RT "Evolution of european human and porcine influenza viruses."
RL Submitted (JAN-2000) to the EMBL/GenBank/DDAJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ252132; CAC81018.1; -;
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63529 MW; 6AA44C84B4DDE68A CRC64;
Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAAGFIENGWEGMIDGWYG 24
Db 346 GLFGAAGFIENGWEGMIDGWYG 368
RESULT 8
Q91MA7
ID Q91MA7 PRELIMINARY; PRT; 566 AA.
AC Q91MA7;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Hemagglutinin.
OS Influenza A virus (A/Hong Kong/1/68 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=108859;
RN [1] _____
RP SEQUENCE FROM N.A.
RC STRAIN=A/Hong Kong/1/68 (H3N2);
RX MEDLINE=21287244; PubMed=11371620;


```

RA Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;
RT "Pattern of mutation in the genome of influenza A virus on adaptation
RT to increased virulence in the mouse lung: Identification of functional
RT themes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AF348176; AAK51718.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63387 MW; 01BB0D465BE158E1 CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
DB 346 GLFGAAGFIENGWEGMIDGWYG 368

RESULT 9
QSDHGO
ID Q9DHGO PRELIMINARY; PRT; 566 AA.
AC Q9DHGO
DT 01-WAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Haemagglutinin precursor.
OS Influenza A virus H3N2.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=41857;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-clone 7a;
RA Mohsin M.A., Morris S.J., Smith H., Sweet C.;
RT "Influenza virus-induced apoptosis: a dual role for viral
RT neuraminidase.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ289703; CAC18525.1; -.
DR HSP; P03437; 2VIU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT SIGNAL 1 16 POTENTIAL.
SQ SEQUENCE 566 AA; 63356 MW; 0BA681929300F72F CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
DB 346 GLFGAAGFIENGWEGMIDGWYG 368

RESULT 10
Q910MS
ID Q910MS PRELIMINARY; PRT; 566 AA.

```

```

AC Q910MS;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hemagglutinin
OS Influenza A virus (A/Hong Kong/1/68 (H3N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=108859;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Hong Kong/1/68 (H3N2);
RX MEDLINE=21287244; PubMed=11371620;
RA Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;
RT "Pattern of mutation in the genome of influenza A virus on adaptation
RT to increased virulence in the mouse lung: Identification of functional
RT themes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AF348179; AAK51721.1; -.
DR EMBL; AF348177; AAK51719.1; -.
DR EMBL; AF348178; AAK51720.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63530 MW; 7CB9F5BFAF6E9F4 CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
DB 346 GLFGAAGFIENGWEGMIDGWYG 368

RESULT 11
Q67126
ID Q67126 PRELIMINARY; PRT; 566 AA.
AC Q67126;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hemagglutinin.
GN HA.
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Seal/MA/3984/92;
RX MEDLINE=95146951; PubMed=7844533;
RA Callan R.J., Early G., Kida H., Hinehaw V.S.;
RT "The appearance of H3 influenza viruses in seals.";
RJ Gen. Virol. 76:199-203(1995).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; L32024; AAA64228.1; -.
DR HSP; P03437; 2VIU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.

```

```
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63441 MW; 590576CB4CE7D08 CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
DB 346 GLFGAAGFIENGWEGMIDGWYG 368

RESULT 12
Q9DXE3 PRELIMINARY; PRT; 301 AA.
AC Q9DXE3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Hemagglutinin (Fragment).
GN HAI.
OS Influenza A virus (A/Shorebird/Taiwan/31-4/99).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=140665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Shorebird/Taiwan/31-4/99;
RA Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee L.H., Shieh H.K.;
RT "Identification and subtyping of avian influenza virus by reverse
transcription-polymerase chain reaction."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HAI AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AF311750; RAQ33016.1; -.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 301
FT CHAIN 301 301
FT CHAIN 301 301
SQ SEQUENCE 301 AA; 32701 MW; 62A403758B764D57 CRC64;

Query Match 93.0%; Score 133; DB 12; Length 301;
Best Local Similarity 95.7%; Pred. No. 6.2e-10;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
DB 250 GLFGAAGFIENGWEGMIDGWYG 272

RESULT 13
Q82499 PRELIMINARY; PRT; 550 AA.
AC Q82499;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Hemagglutinin HAI and HA2 (Fragment).
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Philippines/2/82;
RA Nakajima S., Takeuchi Y., Nakajima K.;
RT "Location on the evolutionary tree of influenza H3 haemagglutinin
genes of Japanese strains isolated during the 1985-6 season."
RL Epidemiol. Infect. 100:301-310(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HAI AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; U08858; AAA18781.1; -.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 550
FT CHAIN 330 550
FT CHAIN 330 550
SQ SEQUENCE 550 AA; 61802 MW; 114413B1CE5A1F6A CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 1.6e-09;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
RT oligosaccharide from the hemagglutinin molecule."
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HAI AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; U08859; AAA18782.1; -.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 550
FT CHAIN 330 550
FT CHAIN 330 550
SQ SEQUENCE 550 AA; 61772 MW; 50BD62B6FE11FD8 CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 1.6e-09;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
DB 330 GLFGAAGFIENGWEGMIDGWYG 352

RESULT 14
Q82498 PRELIMINARY; PRT; 550 AA.
AC Q82498;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Hemagglutinins HAI and HA2 (Fragment).
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
OX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Philippines/2/82;
RA Hartley C.A., Ward A.C., Anders E.M.;
RT "Virulence of influenza virus for mice is associated with loss of
oligosaccharide from the hemagglutinin molecule."
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Philippines/2/82;
RX MEDLINE=88185444; PubMed=3356226;
RA Nakajima S., Takeuchi Y., Nakajima K.;
RT "Location on the evolutionary tree of influenza H3 haemagglutinin
genes of Japanese strains isolated during the 1985-6 season."
RL Epidemiol. Infect. 100:301-310(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HAI AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; U08858; AAA18781.1; -.
DR HSSP; P03437; 2VIU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 550
FT CHAIN 330 550
FT CHAIN 330 550
SQ SEQUENCE 550 AA; 61802 MW; 114413B1CE5A1F6A CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;
```

Best Local Similarity 95.7%; Pred. No. 1.6e-09; Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
|:|||||
Db 330 GIFGAAGFIENGWEGMIDGWYG 352

RESULT 15

Q82753 PRELIMINARY; PRT; 550 AA.
AC Q82753;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Haemagglutinin (Fragment).
OS Influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC unclassified Orthomyxoviridae.
OX NCBI_TaxID=11309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Philippines/2/82/BS/ML10;
RX MEDLINE=97300854; PubMed=9155874;
RA Hartley C.A.; Reading P.C.; Ward A.C.; Anders E.M.;
RT "Changes in the hemagglutinin molecule of influenza type A (H3N2)
RL Arch. Virol. 142:75-88(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Philippines/2/82/BS/ML10;
RX MEDLINE=97456249; PubMed=9311563;
RA Ward A.C.;
RT "Virulence of influenza A virus for mouse lung."
RL Virus Genes 14:187-194(1997).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; U08905; AAC79579.1; -.
DR HSP; P03437; 2VIU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1
FT CHAIN 1 328 HAEMAGGLUTININ HA1.
FT CHAIN 330 550 HAEMAGGLUTININ HA2.
SQ SEQUENCE 550 AA; 61745 MW; 692A49DE678AC4BC CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 1.6e-09;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
|:|||||
Db 330 GIFGAAGFIENGWEGMIDGWYG 352

Search completed: January 30, 2004, 00:24:41
Job time : 26.7887 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	103	49.3	248	1	HSUR1P	histone H1, gonada
2	102	48.8	517	2	T49173	hypothetical prote
3	101.5	48.6	442	2	T39683	zucrin-like protei
4	100	47.8	215	2	T52523	nucleoporin p62 ho
5	100	47.8	380	2	T46395	hypothetical prote
6	98	46.9	206	2	S09388	histone H1 - sea u
7	96	45.9	433	2	S25194	zucrin - yeast (Sa
8	95	45.5	166	2	T18513	hypothetical prote
9	91	43.5	229	2	TJ9219	nuclear protein SR
10	90.5	43.3	383	2	AB6315	F2H15.19 protein -
11	90	43.1	153	2	S95991	histone H2B (clone
12	90	43.1	483	2	F71619	hypothetical prote
13	90	43.1	529	2	T50609	hypothetical prote
14	89	42.6	208	2	T23778	hypothetical prote
15	89	42.6	265	2	S13113	histone H1.1 - Cae
16	88	42.1	392	2	T15755	cgr-4 protein - C
17	88	42.1	409	2	T18726	hypothetical prote
18	87	41.6	107	2	C86477	protein F1504.29 [
19	87	41.6	565	2	T47775	hypothetical prote
20	86	41.1	347	2	E83525	TOLA protein PA097
21	86	41.1	455	2	A87913	protein B0205.10 [
22	85.5	40.9	1280	2	G96796	hypothetical prote
23	85	40.7	103	2	D64363	ribosomal protein
24	85	40.7	152	2	S59587	histone H2B (clone
25	85	40.7	241	2	JN0748	histone H1-II - Vo
26	85	40.7	560	2	T06377	SAR DNA-binding pr
27	85	40.7	849	2	T01286	probable RNA-bindi
28	84.5	40.4	228	2	T36379	probable peptidase
29	84	40.2	409	2	T24543	hypothetical prote

```

Db      35 KKKKKKKKKKKKKKKKKKKKK 54

RESULT 5
T46395
hypothetical protein DKFZp434I1120.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46395
R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23031
A:Accession: T46395
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-380 <AA>
A:Cross-references: ENBL:AL137556
A:Experimental source: adult testis; clone DKFZp434I1120
C:Genetics:
A>Note: DKFZp434I1120.1

Query Match      47.8%; Score 100; DB 2; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps

QY      26 KKKKKKKKKKKKKKKKKKK 45
        |||||
Db      355 KKKKKKKKKKKKKKKKKKK 374

RESULT 6
S09388
histone H1 - sea urchin (Parechinus angulosus)
C:Species: Parechinus angulosus (angulate urchin)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C:Accession: S09388
R:Hill, C.S.; Martin, S.R.; Thomas, J.O.
EMBO J. 8, 2591-2599, 1989
A>Title: A stable alpha-helical element in the carboxy-terminal domain of fr
A:Reference number: S09388; MUID:90060019; PMID:2583125

```

```
A;Accession: S09368
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-206 <HIL>
C;Superfamily: histone H1
C;Keywords: Chromosomal protein

Query Match          46.9%; Score 98; DB 2; Length 206;
Best Local Similarity 55.8%; Pred. No. 0.18;
Matches 24; Conservative 4; Mismatches 15; Indels 0; Gaps

QY      2 EAAAEAEAAAAEAAAEEAAAAKKKKKKKKKKKKKKKKK 44
       : ||| : ||| : ||| : ||| : ||| : |||
Db      140 KAAAKRKKAALAKKAAAKAAATKAKAKKPCKTKAAKAKK 182

RESULT 7
S25194
zuotin - yeast (Saccharomyces cerevisiae)
N;Alternate names: probable Z-DNA-binding protein; protein G9554; protein YC
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence revision 28-May-1993 #text_change 21-Jul-2000
C;Accession: S25194; S64620; S19066
R;Zhang, S.; Lockshin, C.; Herbert, A.; Winter, E.; Rich, A.
EMBO J. 11, 3787-3796, 1992
A;Title: Zuotin, a putative Z-DNA binding protein in Saccharomyces cerevisiae
A;Reference number: S25194; MUID:93010971; PMID:1396572
A;Accession: S25194
A;Molecule type: DNA
A;Residues: 1-433 <ZH>
A;Cross-references: EMBL:X63612; NID:g4836; PIDN:CMA45156.1; PID:g4837
A;Note: part of this sequence, including the amino end of the mature protein
R;Voet, M.; Volckaert, G.
```



```

RX MEDLINE=96120862; PubMed=8590479;
RA Fabry S., Mueller K., Lindauer A., Park P.B., Cornelius T.,
RA Schmitt R.;
RT "The organization structure and regulatory elements of Chlamydomonas
RL histone genes reveal features linking plant and animal genes.";
RL Curr. Genet. 28:333-345(1995).
CC -1- SUBUNIT: The nucleosome is an octamer containing two molecules
CC each of H2A, H2B, H3 and H4. The octamer wraps approximately 146
CC BP of DNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the histone H2B family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC tion between the Swiss Institute of Bioinformatics and the EMBL outsta-
CC the European Bioinformatics Institute. There are no restrictions of
CC use by non-profit institutions as long as its content is in n-
CC modified and this statement is not removed. Usage by and for commu-
CC entities requires a license agreement (See http://www.isb-sib.ch/ann-
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U16726; AAA98454.1; --
CC PIR; S59591; S59591.
CC InterPro; IPR004822; Histone_core.
CC InterPro; IPR000559; Histone_H2B.
CC Pfam; PF00125; histone; 1.
CC PRINTS; PR00621; HISTONEH2B.
CC ProDom; PD000497; Histone_H2B; 1.
CC SMART; SM00427; H2B; 1.
CC PROSITE; PS00357; HISTONE_H2B; 1.
CC Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding;
KW Multigene family.
KQ
SQ SEQUENCE 153 AA; 16587 MW; B09C1ED804E1F485 CRC64;

Query Match 43.1%; Score 90; DB 1; Length 153;
Best Local Similarity 51.0%; Pred. No. 0.14; 16; Indels 6; Gaps
Matches 25; Conservative 2; Mismatches 16;

QY 3 AAAAAAEEEEAAAAAEEEEAAAAKKKKKKKKKKK-----KKKKKKKKK 45
DB 12 ABAGAAPAKAAKPAKAAKAAKKAKPKGDKKKKKK 60

RESULT 6
H11 CAEEL
ID ID H11 CAEEL STANDARD; PRT; 207 AA.
AC AC P1071;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1.1.
GE GN 24.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90204554; PubMed=1969492;
RA Sanicola M., Ward S., Childs G., Emmons S.W.;
RL "Identification of a Caenorhabditis elegans histone H1 gene family.
RT Characterization of a family member containing an intron and encoding
RT a poly(A)+ mRNA.";
RL J. Mol. Biol. 212:259-268(1990).
RN [2]
RP SEQUENCE.
RC STRAIN=Bristol N2;
RX MEDLINE=89076229; PubMed=3202838;
RA Vanfleteren J.R., van Bun S.M., van Beeumen J.J.;
RL "The primary structure of the major isoform (H1.1) of histone H1 from
RT the nematode Caenorhabditis elegans.";
RL Biochem. J. 255:647-652(1988).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.

```



```
Matches 19; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
QY 11 AAAAAAAAAAAAAKKKKKKKKKKKKKKKKKKKKKKKK 45
DB 472 AVHEQLAALSQAPVNFPPKKKKKKKKKKKKKKKK 506

RESULT 13
RL12_PYRAB STANDARD; PRT; 107 AA.
AC Q9UXS6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 50S ribosomal protein L12p.
GN RPL12P OR PYRAB17820 OR PAB1168.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RX PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
CC ACCURATE TRANSLATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A248288; CAB50687.1; -.
CC PIR; A75031; A75031.
CC InterPro; IPR001813; 60s ribosomal.
CC Pfam; PF00428; 60s ribosomal; 1.
CC Ribosomal protein; Complete proteome.
CC SEQUENCE 107 AA; 11300 MW; CD4239930BAD4ED CRC64;
Query Match 39.7%; Score 83; DB 1; Length 107;
Best Local Similarity 44.2%; Pred. No. 0.39;
Matches 19; Conservative 12; Mismatches 12; Indels 0; Gaps 0;
QY 2 EAAAAAAAAAAAAAAAAAAAAAAAAKKKKKKKKKKKKKKKK 44
DB 55 EKAMPVAVAAAPAAPAPAEAPAEKKEEKEEKEEVEE 97

RESULT 14
YD33_YEAST STANDARD; PRT; 320 AA.
AC Q12117;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 36.2 kDa protein in RAD28-LYS14 intergenic region.
GN YDR033W OR YD9673.03.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JM105;
RX MEDLINE=90078104; PubMed=2687247;
RA Levengood S.K., Webster R.E.;
RT "Nucleotide sequences of the tola and tola genes and localization of
RT their products, components of a multistep translocation system in
RT Escherichia coli.";
RL J. Bacteriol. 171:6600-6609(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
```

```
RP SEQUENCE FROM N.A.
RA Arnold W., Becker A., Jaeger W., Kuester H., Nuesbaumer B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ARCHAEAL OPSIN FAMILY. HSP30
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z74329; CAA98855.1; -.
CC EMBL; Z68196; CAA92370.1; -.
CC PIR; S61586; S61586.
CC SGD; S0002440; YDR033W.
CC GO; GO:0005886; C:plasma membrane; IDA.
CC InterPro; IPR001425; Bac_rhodopsin.
CC Pfam; PF01036; Bac_rhodopsin; 1.
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 35 55 POTENTIAL.
CC FT TRANSMEM 63 83 POTENTIAL.
CC FT TRANSMEM 117 137 POTENTIAL.
CC FT TRANSMEM 142 162 POTENTIAL.
CC FT TRANSMEM 168 188 POTENTIAL.
CC FT TRANSMEM 205 225 POTENTIAL.
CC FT TRANSMEM 239 259 POTENTIAL.
CC FT DOMAIN 300 318 LYS-RICH.
CC FT SEQUENCE 320 AA; 36190 MW; 4311F64EAA209F CRC64;
Query Match 39.7%; Score 83; DB 1; Length 320;
Best Local Similarity 48.8%; Pred. No. 1;
Matches 21; Conservative 4; Mismatches 14; Indels 4; Gaps 1;
QY 7 AAAAAAAAAAAAAA-----AKKKKKKKKKKKKKKKKK 45
DB 277 AEAVMAPKAPVASPRPAATPNLSKDKKKKKKKKKKK 319

RESULT 15
TOLA_ECOLI STANDARD; PRT; 421 AA.
AC P19934;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tola protein.
GN TOLA OR CIM OR EXCC OR LKY OR B0739.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JM105;
RX MEDLINE=90078104; PubMed=2687247;
RA Levengood S.K., Webster R.E.;
RT "Nucleotide sequences of the tola and tola genes and localization of
RT their products, components of a multistep translocation system in
RT Escherichia coli.";
RL J. Bacteriol. 171:6600-6609(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
```

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saio N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Iano M., Horiiuchi T.;
RA "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP DOMAINS.
RX MEDLINE=91296736; PubMed=2068069;
RA Levengood S.K., Beyer W.F. Jr., Webster R.E.;
RT "Tola: a membrane protein involved in colicin uptake contains an
RT extended helical region.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).
RN [5]
RP INTERACTION WITH PORINS.
RX MEDLINE=97133271; PubMed=8978668;
RA Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,
RA Lloubes R.;
RT "Tola central domain interacts with *Escherichia coli* porins.";
RL EMBO J. 15:6408-6415(1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.
RX MEDLINE=99332679; PubMed=10404600;
RA Lubkowski J., Henneke F., Plueckthun A., Wlodawer A.;
RT "Filamentous phage infection: crystal structure of g3p in complex
RT with its coreceptor, the C-terminal domain of Tola.";
RL Structure 7:711-722(1999).
CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A
CC COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE
CC COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL
CC BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION
CC OF BACTERIOPHAGE DNA.
CC -!- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE
CC AND LAMB.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M28232; AA24683.1; -;
CC EMBL; AE000177; AAC73833.1; -;
CC EMBL; D90713; BA335405.1; -;
CC PIR; JVO057; JVO057.
CC PDB; 1TOL; 20-MAY-99.
CC EcGene; Egl1007; tola.
KW Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW Repeat; Inner membrane; 3D-structure; Complete proteome.
FT DOMAIN 1 13
FT TRANSMEM 14 34
FT DOMAIN 35 421 PERIPLASMIC (POTENTIAL).
FT DOMAIN 48 310 DOMAIN II (ALPHA-HELICAL).
FT DOMAIN 311 421 DOMAIN III (FUNCTIONAL).
FT DOMAIN 224 278 10 X TANDEM REPEATS OF [ED]-K(1,2)-
FT A(2,4).
FT HELIX 335 349

FT TURN 350 351
FT TURN 353 354
FT HELIX 355 358
FT TURN 359 360
FT STRAND 363 369
FT TURN 371 372
FT STRAND 375 383
FT HELIX 385 397
FT HELIX 406 412
FT TURN 413 414
FT STRAND 416 421
SQ SEQUENCE 421 AA; 8B2F52B4B97C655E CRC64;
Query Match 39.7%; Score 83; DB 1; Length 421;
Best Local Similarity 47.1%; Pred. No. 1.3;
Matches 24; Conservative 5; Mismatches 14; Indels 8; Gaps 1;
QY 3 AAAAAEAAAAEAAA-----AAEAAAAA KKKKKKKKKKKKKKKKK 45
DB 176 AKAAAEAKKAEAAAAAALKKKKAEAAEAAAEAKKAAATEAAEAKAEKK 226

Search completed: January 30, 2004, 00:20:46
Job time : 11.1408 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:52 ; Search time 46.4789 Seconds
(without alignments)
249.842 Million cell updates/sec

Title: US-09-461-684C-4

Perfect score: 209

Sequence: 1 CEAAAERAAAAAEEA...KKKKKKKKKKKKKKKKKKKK 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	56.5	1038	5 Q8MQW9	Q8MQW9 drosophila
2	118	56.5	2347	5 Q8INH9	Q8INH9 drosophila
3	118	56.5	2451	5 Q9VG05	Q9VG05 drosophila
4	107	51.2	791	5 Q8T2U7	Q8T2U7 dictyosteli
5	105	50.2	129	11 Q35807	Q35807 rattus norv
6	105	50.2	720	4 Q9H6Q7	Q9H6Q7 homo sapien
7	104	49.8	658	11 Q8CGI8	Q8CGI8 mus musculu
8	103	49.3	113	10 Q8LQP6	Q8LQP6 oryza sativ
9	102	48.8	467	10 Q8LL82	Q8LL82 englena gra
10	102	48.8	517	10 Q8LXR2	Q8LXR2 arabidopsis
11	101.5	48.6	354	3 O14347	O14347 schizosacch
12	101	48.3	531	6 Q9SLV6	Q9SLV6 macaca fasc
13	100.5	48.1	80	10 Q8S7D3	Q8S7D3 oryza sativ
14	100	47.8	55	4 Q8N6F0	Q8N6F0 homo sapien
15	100	47.8	128	3 Q9P529	Q9P529 neurospora
16	100	47.8	168	4 Q9H5V6	Q9H5V6 homo sapien

17	100	47.8	206	5 Q8I247	Q8I247 plasmodium
18	100	47.8	215	11 Q64075	Q64075 rattus sp.
19	100	47.8	260	10 Q9LG29	Q9LG29 arabidopsis
20	100	47.8	380	4 Q9NT34	Q9NT34 homo sapien
21	100	47.8	515	5 Q8SWR7	Q8SWR7 drosophila
22	99	47.4	686	4 Q9NXF0	Q9NXF0 homo sapien
23	97	46.4	476	10 Q8RVV4	Q8RVV4 lycopersico
24	97	46.4	667	4 Q9HC48	Q9HC48 homo sapien
25	95	45.5	257	4 Q9H5Y3	Q9H5Y3 homo sapien
26	95	45.5	489	5 Q8T3H4	Q8T3H4 drosophila
27	93	44.5	332	4 Q8IZU1	Q8IZU1 homo sapien
28	92.5	44.3	372	2 Q8WXX1	Q8WXX1 pseudomonas
29	92	44.0	351	16 Q8FQZ8	Q8FQZ8 corynebacte
30	92	44.0	433	5 Q8IJJ5	Q8IJJ5 plasmodium
31	91	43.5	229	11 Q9JMG3	Q9JMG3 mus musculu
32	91	43.5	263	4 Q8BU76	Q8BU76 homo sapien
33	91	43.5	263	4 Q8IZH3	Q8IZH3 homo sapien
34	91	43.5	556	10 Q8S405	Q8S405 hordeum vul
35	91	43.5	3351	5 Q8IBK4	Q8IBK4 plasmodium
36	90.5	43.3	368	10 Q8LCH0	Q8LCH0 arabidopsis
37	90.5	43.3	368	10 Q944L9	Q944L9 arabidopsis
38	90.5	43.3	383	10 Q9LMT3	Q9LMT3 arabidopsis
39	90	43.1	39	5 Q8MSJ8	Q8MSJ8 drosophila
40	90	43.1	166	5 Q928I1	Q928I1 parechinus
41	90	43.1	180	5 Q25636	Q25636 parechinus
42	90	43.1	529	4 Q9NFX4	Q9NFX4 homo sapien
43	90	43.1	628	4 Q9H5M5	Q9H5M5 homo sapien
44	90	43.1	749	5 Q967D9	Q967D9 drosophila
45	90	43.1	1062	5 Q960C4	Q960C4 drosophila

ALIGNMENTS

RESULT 1

Q8MQW9 PRELIMINARY; PRT; 1038 AA.
ID Q8MQW9
AC Q8MQW9; 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE SD05989p (Fragment).
GN CG7518
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY12252; AAM52764.1; -
DR FlyBase; FBgn0038108; CG7518.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON TER 1
SQ SEQUENCE 1038 AA; 109059 MW; 80C935A2C6D8A276 CRC64;

Query Match 56.5%; Score 118; DB 5; Length 1038;

Best Local Similarity 75.7%; Pred. No. 0.01;

Matches 28; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 AA 39

DB 112 AA 148


```

RESULT 2
ID Q8INH9 PRELIMINARY; PRT; 2347 AA.
AC Q8INH9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG7518-PB.
GN CG7518.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fogle C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodruff, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang K., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX CELNIKER S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jallali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."

```

```

Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Miara S., Croebly M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnik S.E.,
RA Klomp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Seale S.M., Smith E., Shu S., Smutniak P., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Flybase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003698; AAN14338.1;
SQ SEQUENCE 2347 AA; 257013 MW; 23BF5FC5FFCABA64 CRC64;
Query Match 56.58; Score 118; DB 5; Length 2347;
Best Local Similarity 75.74; Pred. No. 0.021;
Matches 28; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 3 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 39
DB 1374 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1410

```

```

RESULT 3
Q9VG05 PRELIMINARY; PRT; 2451 AA.
ID Q9VG05
AC Q9VG05;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG7518 protein.
GN CG7518.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fogle C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodruff, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang K., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX CELNIKER S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jallali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."

```

```

DE MICROVASCULAR endothelial differentiation protein 2.
MDG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Epididymis;
RX MEDLINE=98172708; PubMed=9511718;
RA Proels F., Loser B., Marx M.;
RT "Differential expression of osteopontin, PC4, and CRC5, a novel mRNA
RT species, during in vitro angiogenesis.";
RL Exp. Cell Res. 239:1-10(1998).
DR EMBL; Y08769; CAA70022.1; -.
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00069; pkinase; 1.
DR PRODOM; PD000001; Prot_kinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 129 AA; 15080 MW; 38102272BBE2EDB4 CRC64;

Query Match 50.2%; Score 105; DB 11; Length 129;
Best Local Similarity 95.5%; Pred. No. 0.02;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAKKKKKKKKKKKKKKKKKKKK 45
DB 83 ASKKKKKKKKKKKKKKKKKKK 104
|:|||||||
|:|||||||

RESULT 6
ID Q9H6Q7 PRELIMINARY; PRT; 720 AA.
AC Q9H6Q7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ21979 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025632; BAB15196.1; -.
DR Hypothetical protein.
RW NON_TER 720
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Query Match 50.2%; Score 105; DB 4; Length 720;
Best Local Similarity 61.8%; Pred. No. 0.095;
Matches 21; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 12 AAEAAAAAFAAAAKKKKKKKKKKKKKKKKKKKK 45
DB 678 AKKSITNSDIVSISKKKKKKKKKKKKKKKKKKKK 711
|:::|
|:|||||||

RESULT 7
ID Q8CGI8 PRELIMINARY; PRT; 658 AA.
AC Q8CGI8;
DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
```


RESULT 13
Q8S7D3
ID ID Q8S7D3 PRELIMINARY; PRT; 80 AA.
AC Q8S7D3;
DT 01-JUN-2002 (T=EMBLrel. 21, Created)

RESULT 15
Q9P529
ID Q9P529 PRELIMINARY; PRT; 128 AA.
AC Q9P529; AC
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 15.2 kDa protein.
GN B24H17.160.
OS Neurospora crassa.
OC Eukaryota; Fungi;
OC Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL356815; CAB92638.2; --
 KW Hypothetical protein.
 SQ SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFB70765 CRC64;
 Query Match 47.8%; Score 100; DB 3; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.055;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 26 KKKKKKKKKKKKKKKKKKK 45
 Db 71 KKKKKKKKKKKKKKKKKKK 90

Search completed: January 30, 2004, 00:24:41
 Job time : 46.4789 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 00:13:12 ; Search time 18.7981 Seconds
(without alignments)
225.098 Million cell updates/sec

Title: US-09-461-684C-5

Perfect score: 243

Sequence: 1 GLFGAIAAGFIENGWEGMID.....KKKKKKKKKKKKKKKKKKKK 44

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	55.1	550	1 HMIVS2	hemagglutinin prec
2	134	55.1	550	1 HMIVS3	hemagglutinin prec
3	134	55.1	550	1 HMIV77	hemagglutinin prec
4	134	55.1	550	1 HMIV80	hemagglutinin prec
5	134	55.1	550	1 HMIV33	hemagglutinin prec
6	134	55.1	550	1 HMIV89	hemagglutinin prec
7	134	55.1	550	1 HMIV21	hemagglutinin prec
8	134	55.1	550	1 HMIV98	hemagglutinin prec
9	134	55.1	550	1 HMIV15	hemagglutinin prec
10	134	55.1	550	2 JQ1153	hemagglutinin prec
11	134	55.1	550	2 JQ1154	hemagglutinin prec
12	134	55.1	550	2 JQ1155	hemagglutinin prec
13	134	55.1	566	1 HMIVH	hemagglutinin prec
14	134	55.1	566	1 HMIVHA	hemagglutinin prec
15	134	55.1	566	1 HMIVHM	hemagglutinin prec
16	134	55.1	566	1 HMIVDU	hemagglutinin prec
17	134	55.1	570	1 A45591	hemagglutinin prec
18	134	55.1	570	2 S22013	hemagglutinin prec
19	133	54.7	561	1 HMIV49	hemagglutinin prec
20	133	54.7	561	1 HMIV84	hemagglutinin prec
21	132	54.3	565	1 HMIV81	hemagglutinin prec
22	132	54.3	565	1 HMIV83	hemagglutinin prec
23	132	54.3	566	1 HMIV6	hemagglutinin prec
24	132	54.3	567	1 HMIVV	hemagglutinin prec
25	131	53.9	362	2 S38637	hemagglutinin - in
26	131	53.9	550	1 HMIV86	hemagglutinin prec
27	131	53.9	560	1 HMIVT7	hemagglutinin prec
28	131	53.9	565	1 HMIV82	hemagglutinin prec
29	131	53.9	565	1 HMIV84	hemagglutinin prec

30 131 53.9 565 1 HMIVS5 hemagglutinin prec
31 131 53.9 565 1 HMIVS6 hemagglutinin prec
32 131 53.9 565 1 HMIVS7 hemagglutinin prec
33 131 53.9 565 1 HMIVS8 hemagglutinin prec
34 131 53.9 565 1 HMIVS9 hemagglutinin prec
35 131 53.9 565 1 HMIVS1 hemagglutinin prec
36 131 53.9 565 1 HMIVS2 hemagglutinin prec
37 131 53.9 565 2 S33703 hemagglutinin - in
38 131 53.9 570 2 S33703 hemagglutinin - in
39 131 53.9 570 2 S22014 hemagglutinin prec
40 131 53.9 570 2 S22015 hemagglutinin prec
41 131 53.9 570 2 S22016 hemagglutinin prec
42 131 53.9 570 2 S22017 hemagglutinin prec
43 131 53.9 570 2 S22020 hemagglutinin prec
44 131 53.9 570 2 S22021 hemagglutinin prec
45 131 53.9 570 2 S22029 hemagglutinin prec

ALIGNMENTS

RESULT 1

HMIVS2
hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)
C:Species: influenza A virus
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: A29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Viology 162, 160-166, 1988

A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A:Reference number: A94370; MUID:88101364; PMID:3336940
A:Accession: A29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>

A:Cross-references: GB:M19056; MID:g324208
A:Note: the sequence in Genbank entry FLAHPA, release 106, (PID:g324209) differs from the
C:Genetics:
C:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
P:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F:520-536/Domain: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 2

HMIVS3
hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)
C:Species: influenza A virus
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: B29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Viology 162, 160-166, 1988

A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A:Reference number: A94370; MUID:88101364; PMID:3336940
A:Accession: B29971
A:Molecule type: Genomic RNA
A:Residues: 1-550 <KID>

A:Cross-references: GB:M19057; MID:g324210
A:Note: the sequence in Genbank entry FLAHPB, release 106, (PID:g324211) differs from the
C:Genetics:
C:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Product: hemagglutinin HA2 #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||
DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 3
HMIV77
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A27813
R:Kida, H.; Kawaoka, Y.; Naeye, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; PMID:87265458; PMID:2440178
A:Accession: A27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16737; NID:G324081; PIDN:AAA43143.1; PID:G324082
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Product: hemagglutinin HA2 #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||
DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 4
HMIV80
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C:Accession: B27813
R:Kida, H.; Kawaoka, Y.; Naeye, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; PMID:87265458; PMID:2440178
A:Accession: B27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16738; NID:G324083
A:Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1

C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Product: hemagglutinin HA2 #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||
DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 5
HMIV33
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: C27813
R:Kida, H.; Kawaoka, Y.; Naeye, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; PMID:87265458; PMID:2440178
A:Accession: C27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16739; NID:G324085; PIDN:AAA43145.1; PID:G324086
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Product: hemagglutinin HA2 #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||
DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 6
HMIV89
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: D27813
R:Kida, H.; Kawaoka, Y.; Naeye, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; PMID:87265458; PMID:2440178
A:Accession: D27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16740; NID:G324087; PIDN:AAA43146.1; PID:G324088
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Product: hemagglutinin HA2 #status predicted <TM1>

F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 7

HMIv21
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: E27813
R;Kida, H.; Kawakoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: E27813
A;Molecule type: Genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16741; NID:g324089
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane site: carbohydrate (Asn) (covalent) #status predicted <TM1>
F;7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 8

HMIv98
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: F27813
R;Kida, H.; Kawakoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: F27813
A;Molecule type: Genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16742; NID:g324091
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane site: carbohydrate (Asn) (covalent) #status predicted <TM1>
F;7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 9

HMIv15
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: G27813
R;Kida, H.; Kawakoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: G27813
A;Molecule type: Genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16743; NID:g324093; PIDN:AAA43149.1; PID:g324094
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane site: carbohydrate (Asn) (covalent) #status predicted <TM1>
F;7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 10

JQ1153
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1153
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1153
A;Molecule type: Genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00929; NID:g221279; PIDN:BA00769.1; PID:g221280
A;Note: the authors translated the codon GGG for residue 218 as Glu
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
|||||
Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 11

QJ1154

hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000

C:Accession: JQ1154

R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.

J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3

A:Reference number: JQ1153; MUID:91341491; PMID:1875195

A:Accession: JQ1154

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D09030; NID:9221273; PIDN:BAA00770.1; PID:9221274

A:Note: the authors translated the codon GGG for residue 218 as Glu

A:Note: residues 528-532 are not shown in this publication

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;

Best Local Similarity 100.0%; Pred. No. 3.1e-06;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24

Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 12

QJ1155

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000

C:Accession: JQ1155

R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.

J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3

A:Reference number: JQ1153; MUID:91341491; PMID:1875195

A:Accession: JQ1155

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D09031; NID:9221277; PIDN:BAA00771.1; PID:9221278

A:Note: the authors translated the codon GGG for residue 218 as Glu, GCC for residue 538

A:Note: residues 528-532 are not shown in this publication

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;

Best Local Similarity 100.0%; Pred. No. 3.1e-06;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24

Db 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 13

HM1VH

hemagglutinin precursor - influenza A virus

C:Species: Influenza A virus

C>Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 22-Oct-1999

C:Accession: A93705; A93233; A04051; A93231; A94441

R:Both, G.W.; Sleight, M.J.

Nucleic Acids Res. 8, 2561-2575, 1980

A:Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza

A:Reference number: A93705; MUID:81053698; PMID:6253883

A:Accession: A93705

A:Molecule type: genomic RNA

A:Residues: 1-566 <BOT>

A:Cross-references: GB:V01103

A:Experimental source: strain A/NT/60/68/29C

A:Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/NT/6

R:Dopheide, T.A.; Ward, C.W.

FEBS Lett. 110, 181-193, 1980

A:Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.

A:Reference number: A91276; MUID:80179105; PMID:6769586

A:Contents: annotation; disulfide bonds

R:Gething, M.J.; Bye, J.; Skehel, J.; Waterfield, M.

Nature 287, 301-306, 1980

A:Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from

A:Reference number: A93233; MUID:81030852; PMID:7421990

A:Accession: A93233

A:Molecule type: genomic RNA

A:Residues: 1-24, 'S', '26', 'D', '28-159, 'G', '161-197, 'I', '199-241, 'L', '243-249 <GET>

A:Experimental source: strain X-31[H3]

C:Superfamily: influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:536-552/Domain: transmembrane #status predicted <TM1>

F:30-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status experimental

F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;

Best Local Similarity 100.0%; Pred. No. 3.2e-06;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24

Db 346 GLFGAIAAGFIENGWEGMIDGWYG 368

RESULT 14

HM1VHA

hemagglutinin precursor - influenza A virus (strain A/Aichi/2/68)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus

C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999

C:Accession: A93231; A04051

R:Verhoeven, M.; Fang, R.; Min Jou, W.; Devos, R.; Huylebroeck, D.; Saman, E.; Fiers, W.

Nature 286, 771-776, 1980

A:Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A.

A:Reference number: A93231; MUID:80254693; PMID:7402351

A:Accession: A93231

A:Molecule type: genomic RNA

A:Residues: 1-566 <VER>

A:Cross-references: GB:J02090; NID:9324131; PIDN:AAA43178.1; PID:9324132

C:Superfamily: influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;

Best Local Similarity 100.0%; Pred. No. 3.2e-06;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24

Db 346 GLFGAIAAGFIENGWEGMIDGWYG 368

RESULT 15

HM1VHM
hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 31-Mar-2000
C;Accession: A94441; A04051
R;Stleigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.
in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, B
A;Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of c
A;Reference number: A94441
A;Accession: A94441
A;Molecule type: genomic RNA
A;Residues: 1-566 <SLE>
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 3.2e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAIAAGFIENGWEGMIDGWYG 24
Db 346 GLFGAIAAGFIENGWEGMIDGWYG 368

Search completed: January 30, 2004, 00:26:22
Job time : 18.7981 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:27 ; Search time 9.91549 Seconds
(without alignments)
208.681 Million cell updates/sec

Title: US-09-461-684C-5

Perfect score: 243

Sequence: 1 CGLFGAIAFIENGWEGMID.....KKKKKKKKKKKKKKKKKKKK 44

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	55.1	550	HEMA_IADH1	P12582 influenza a
2	134	55.1	550	HEMA_IADH2	P12583 influenza a
3	134	55.1	550	HEMA_IADH3	P12584 influenza a
4	134	55.1	550	HEMA_IADH4	P12585 influenza a
5	134	55.1	550	HEMA_IADH5	P12586 influenza a
6	134	55.1	550	HEMA_IADH6	P12587 influenza a
7	134	55.1	550	HEMA_IADH7	P12588 influenza a
8	134	55.1	550	HEMA_IADHK	P43257 influenza a
9	134	55.1	550	HEMA_IADHL	P43258 influenza a
10	134	55.1	550	HEMA_IAGHK	P43260 influenza a
11	134	55.1	550	HEMA_IAZH2	P11133 influenza a
12	134	55.1	550	HEMA_IAZH3	P11134 influenza a
13	134	55.1	566	HEMA_IADH1C	P03437 influenza a
14	134	55.1	566	HEMA_IADH3C	P26134 influenza a
15	134	55.1	566	HEMA_IADH3C	P26134 influenza a
16	134	55.1	566	HEMA_IADH3C	P26134 influenza a
17	134	55.1	566	HEMA_IADH3C	P26134 influenza a
18	134	55.1	566	HEMA_IADH3C	P26134 influenza a
19	134	55.1	566	HEMA_IADH3C	P26134 influenza a
20	134	55.1	570	HEMA_IADH6C	P26094 influenza a
21	134	55.1	570	HEMA_IADH6C	P26101 influenza a
22	134	54.7	561	HEMA_IADH6C	P12581 influenza a
23	133	54.7	561	HEMA_IADH6C	P12581 influenza a
24	132	54.3	565	HEMA_IADH6C	P17000 influenza a
25	132	54.3	565	HEMA_IADH6C	P17002 influenza a
26	132	54.3	566	HEMA_IADH6C	P03440 influenza a
27	132	54.3	566	HEMA_IADH6C	P26139 influenza a
28	132	54.3	567	HEMA_IADH6C	P03435 influenza a
29	131	53.9	550	HEMA_IADH6C	P12589 influenza a
30	131	53.9	560	HEMA_IADH6C	P03458 influenza a
31	131	53.9	565	HEMA_IADH6C	P16994 influenza a
32	131	53.9	565	HEMA_IADH6C	P16995 influenza a
33	131	53.9	565	HEMA_IADH6C	P16999 influenza a

34	131	53.9	565	1	HEMA_IADH7	P16996 influenza a
35	131	53.9	565	1	HEMA_IADH7	P16997 influenza a
36	131	53.9	565	1	HEMA_IADH7	P16998 influenza a
37	131	53.9	565	1	HEMA_IADH7	P16999 influenza a
38	131	53.9	565	1	HEMA_IADH7	P16999 influenza a
39	131	53.9	565	1	HEMA_IADH7	P16999 influenza a
40	131	53.9	565	1	HEMA_IADH7	P16999 influenza a
41	131	53.9	566	1	HEMA_IADH7	P16999 influenza a
42	131	53.9	566	1	HEMA_IADH7	P16999 influenza a
43	131	53.9	570	1	HEMA_IADH7	P16999 influenza a
44	131	53.9	570	1	HEMA_IADH7	P16999 influenza a
45	131	53.9	570	1	HEMA_IADH7	P16999 influenza a

ALIGNMENTS

RESULT 1
HEMA_IADH1
ID HEMA_IADH1 STANDARD; PRT; 550 AA.
AC P12582; Q84021; Q84022;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/5/77).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11357;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawakita Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; M16737; AAA43143.1; -.
CC HSP; P03437; 3HMG.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61705 MW; 7E7ACFE716FC969A CRC64;
Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RN SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawaoka Y., Naeye C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks";
 RL Virology 159:109-119(1987).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M16740; AAA43146.1; -;
 DR HSP; P03437; 2VIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON_TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 178 179 YV -> VI (IN PIR DATA BANK).
 FT CONFLICT 388 388 K -> T (IN PIR DATA BANK).
 SQ SEQUENCE 550 AA; 61664 MW; A1682CF8CBBD9D0 CR664;

 Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 2 GLFGAAGFIENGWEGMIDGWYG 24
 DB 330 GLFGAAGFIENGWEGMIDGWYG 352

 RESULT 5
 HEMA_IADH5 STANDARD; PRT; 550 AA.
 ID HEMA_IADH5
 AC P12586; Q84015; Q84016;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor (Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain) (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11361;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawaoka Y., Naeye C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks";
 RL Virology 159:109-119(1987).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M16741; AAA43147.1; -;
 DR PIR; E27813; HMIV21.
 DR HSP; P03437; 2VIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON_TER 1 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 7 7 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 178 179 YV -> VI (IN PIR DATA BANK).
 FT CONFLICT 388 388 K -> T (IN PIR DATA BANK).
 SQ SEQUENCE 550 AA; 61856 MW; 48401C867A15BF8C CR664;

 Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 2 GLFGAAGFIENGWEGMIDGWYG 24
 DB 330 GLFGAAGFIENGWEGMIDGWYG 352

 RESULT 6
 HEMA_IADH6 STANDARD; PRT; 550 AA.
 ID HEMA_IADH6
 AC P12587; Q84017;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor (Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain) (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11362;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawaoka Y., Naeye C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks";
 RL Virology 159:109-119(1987).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 9
HEMA IADHL
ID HEMA IADHL STANDARD; PRT; 550 AA.
AC P43258;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hong Kong/64/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=45412;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/Hong Kong/68 (H3N2) strain emerged.";
RL J. Gen. Virol. 72:2007-2010(1991).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00931; BAA00771.1; -
CC HSP; P03437; 2VIU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 1 328 HEMAGGLUTININ HA2 CHAIN.
FT CHAIN 330 550 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61718 MW; A351C56789E4BE9A CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 11
HEMA IAZH2
ID HEMA IAZH2 STANDARD; PRT; 550 AA.
AC P11133; Q84019; Q84020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE chain] (Fragment).
GN HA.

RESULT 10
HEMA IAGHK
ID HEMA IAGHK STANDARD; PRT; 550 AA.
AC P43260;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Goose/Hong Kong/10/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=45414;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/Hong Kong/68 (H3N2) strain emerged.";
RL J. Gen. Virol. 72:2007-2010(1991).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00930; BAA00770.1; -
CC HSP; P03437; 2VIU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 1 328 HEMAGGLUTININ HA2 CHAIN.
FT CHAIN 330 550 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61676 MW; 9A1E094DA28BACD2 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAAGFIENGWEGMIDGWYG 24
DB 330 GLFGAIAAGFIENGWEGMIDGWYG 352

RESULT 11
HEMA IAZH2
ID HEMA IAZH2 STANDARD; PRT; 550 AA.
AC P11133; Q84019; Q84020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE chain] (Fragment).
GN HA.

OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11497;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88101364; PubMed=3336940;
 RA Kida H., Shortridge K.F., Webster R.G.;
 RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
 in China";
 RL Virology 162:160-166(1988).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M19057; AAA43212.1; -;
 CC HSPSP; P03437; 2VIU.
 CC InterPro; IPR001364; Hemagglutn.
 CC Pfam; PF00509; Hemagglutinin; 1.
 CC PRINTS; PR00329; HEMAGGLUTN12.
 CC ProDom; PD000225; Hemagglutn; 1.
 CC Hemagglutinin; Envelope protein; Glycoprotein.
 FT CHAIN 1 1 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61437 MW; 1F2A7E758C531CE8 CRC64;
 Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLFGAIAAGFTIENGWEGMIDGWYG 24
 |||||
 Db 330 GLFGAIAAGFTIENGWEGMIDGWYG 352
 RESULT 12
 HEMA_IASH3
 ID HEMA_IASH3 STANDARD; PRT; 550 AA.
 AC P11134; Q84025; Q84026;
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
 chain] (Fragment).
 DE HA.
 OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=11498;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88101364; PubMed=3336940;
 RA Kida H., Shortridge K.F., Webster R.G.;
 RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
 in China";
 RL Virology 162:160-166(1988).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M19056; AAA43211.1; ALT_TERM.
 CC HSPSP; P03437; 2VIU.
 CC InterPro; IPR001364; Hemagglutn.
 CC Pfam; PF00509; Hemagglutinin; 1.
 CC PRINTS; PR00329; HEMAGGLUTN12.
 CC ProDom; PD000225; Hemagglutn; 1.
 CC Hemagglutinin; Envelope protein; Glycoprotein.
 FT CHAIN 1 1 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61580 MW; 991F6D8BC02F24F2 CRC64;
 Query Match 55.1%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLFGAIAAGFTIENGWEGMIDGWYG 24
 |||||
 Db 330 GLFGAIAAGFTIENGWEGMIDGWYG 352
 RESULT 13
 HEMA_IASH3
 ID HEMA_IASH3 STANDARD; PRT; 566 AA.
 AC P03437;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 DE HA.
 OS Influenza A virus (strain A/Aichi/2/68).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=150147;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80254693; PubMed=7402351;
 RA Verhoeven M., Fang R., Min Jou W., Devos R., Huylebroeck D.,
 RA Saman E., Fiers W.;
 RT "Antigenic drift between the haemagglutinin of the Hong Kong
 RT influenza strains A/Aichi/2/68 and A/Victoria/3/75";
 RL Nature 286:771-776(1980).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RX MEDLINE=81123029; PubMed=7464906;
 RA Wilson I.A., Skehel J.J., Wiley D.C.;
 RT "Structure of the haemagglutinin membrane glycoprotein of influenza
 RT virus at 3-A resolution";
 RL Nature 289:366-373(1981).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY.
 RX MEDLINE=88232903; PubMed=3374584;

RA Weis W.I., Brown J.H., Cusack S.C., Paulson J.C., Skehel J.J.,
RA Wiley D.C.;
RT "Structure of the influenza virus haemagglutinin complexed with its
RT receptor, sialic acid.";
RL Nature 333:426-431(1988).
[4]
RN X-RAY CRYSTALLOGRAPHY OF A MUTANT WITH GLY-457.
RX MEDLINE=90107940; PubMed=2295311;
RA Weis W.I., Cusack S.C., Brown J.H., Daniels R.S., Skehel J.J.,
RA Wiley D.C.;
RT "The structure of a membrane fusion mutant of the influenza virus
RT haemagglutinin";
RL EMBO J. 9:17-24(1990).
[5]
RN X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=90230310; PubMed=2329580;
RA Weis W.I., Bruegger A.T., Skehel J.J., Wiley D.C.;
RT "Refinement of the influenza virus haemagglutinin by simulated
RT annealing";
RL J. Mol. Biol. 212:737-761(1990).
[6]
RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=94352388; PubMed=8072525;
RA Bullough P.A., Hughson F.M., Skehel J.J., Wiley D.C.;
RT "Structure of influenza haemagglutinin at the pH of membrane fusion.";
RL Nature 371:37-43(1994).
[7]
RN X-RAY CRYSTALLOGRAPHY (3.25 ANGSTROMS).
RX MEDLINE=98120975; PubMed=9461077;
RA Fleury D., Wharton S.A., Skehel J.J., Knossow M., Bizebard T.;
RT "Antigen distortion allows influenza virus to escape neutralization.";
RL Nat. Struct. Biol. 5:119-123(1998).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
or send an email to license@ebi.ac.uk).
CC
EMBL: J02090; AAA43178.1; --
EMBL: V01085; CAA24269.1; --
FIR: A93231; HMVHA.
PDB: 2HMG; 31-OCT-93.
PDB: 3HMG; 31-OCT-93.
PDB: 4HMG; 31-OCT-93.
PDB: 5HMG; 31-JAN-94.
PDB: 1HGD; 31-JAN-94.
PDB: 1HGE; 31-JAN-94.
PDB: 1HGF; 31-JAN-94.
PDB: 1HGG; 31-JAN-94.
PDB: 1HGH; 31-JAN-94.
PDB: 1HGI; 31-JAN-94.
PDB: 1HGJ; 31-JAN-94.
PDB: 1HJN; 14-FEB-95.
PDB: 2VIR; 29-APR-98.
PDB: 2VIS; 29-APR-98.
PDB: 2VIT; 29-APR-98.
PDB: 2VIU; 29-APR-98.
PDB: 1E08; 29-NOV-00.
PDB: 1HA0; 22-DEC-99.
PDB: 1J8H; 13-MAR-02.
PDB: 1KEN; 24-APR-02.
PDB: 1QFU; 29-DEC-99.
PDB: 1QUL; 05-JAN-00.
DR InterPro: IPR01364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT DISULFID 30 482 INTERCHAIN.
FT DISULFID 68 293
FT DISULFID 80 92
FT DISULFID 113 155
FT DISULFID 297 321
FT DISULFID 489 493
FT CARBOHYD 24 24
FT CARBOHYD 38 38
FT CARBOHYD 54 54
FT CARBOHYD 97 97
FT CARBOHYD 181 181
FT CARBOHYD 301 301
FT CARBOHYD 499 499
FT STRAND 27 35
FT STRAND 40 42
FT STRAND 50 52
FT STRAND 55 57
FT STRAND 59 60
FT STRAND 67 70
FT STRAND 74 76
FT TURN 78 79
FT HELIX 82 87
FT TURN 88 88
FT HELIX 90 95
FT TURN 96 97
FT STRAND 99 99
FT STRAND 102 105
FT TURN 107 108
FT STRAND 116 117
FT TURN 119 120
FT HELIX 121 131
FT STRAND 133 133
FT STRAND 136 138
FT TURN 144 145
FT STRAND 146 147
FT STRAND 152 157
FT TURN 158 159
FT STRAND 160 162
FT TURN 165 166
FT STRAND 167 169
FT TURN 171 173
FT TURN 174 175
FT STRAND 176 176
FT STRAND 180 185
FT STRAND 192 200
FT HELIX 204 211
FT STRAND 217 221
FT STRAND 226 229
FT STRAND 239 239
FT TURN 240 241
FT STRAND 242 242
FT STRAND 245 253
FT TURN 255 256
FT STRAND 258 265
FT STRAND 267 270
FT STRAND 272 275
FT STRAND 282 285
FT STRAND 290 294
FT STRAND 297 299
FT TURN 300 301
FT STRAND 302 304
FT STRAND 310 311
FT STRAND 318 320
FT STRAND 323 324
FT STRAND 331 333
FT STRAND 337 337
FT TURN 347 348

```
FT TURN 350 350
FT STRAND 351 351
FT TURN 352 354
FT STRAND 353 355
FT TURN 354 355
FT STRAND 355 360
FT TURN 356 361
FT STRAND 357 373
FT TURN 358 373
FT STRAND 359 375
FT TURN 360 382
FT STRAND 361 382
FT TURN 362 400
FT STRAND 363 401
FT TURN 364 401
FT STRAND 365 406
FT TURN 366 407
FT STRAND 367 421
FT TURN 368 421
FT STRAND 369 471
FT TURN 370 474
FT STRAND 371 477
FT TURN 372 485
FT STRAND 373 485
FT TURN 374 491
FT STRAND 375 498
FT TURN 376 499
FT STRAND 377 500
FT TURN 378 504
FT STRAND 379 515
FT TURN 380 518
FT STRAND 381 519
FT TURN 382 566 AA; 63415 MW; E395659C23CAFECA CRC64;
SQ SEQUENCE 566 AA; 63415 MW; E395659C23CAFECA CRC64;

Query Match 55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAAGFIENGWEGMIDWYG 24
| | | | | | | | | | | | | | | | | | | | | |
Db 346 GLFGAAGFIENGWEGMIDWYG 368

RESULT 14
HEMA_IADA3
ID HEMA_IADA3 STANDARD; PRT; 566 AA.
AC P26134;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain]
GN HA.
OS Influenza A virus (strain A/Duck/Alberts/78/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11348;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Schell M., Katz J., Kawaoka Y., Naeye C., Gorman O.,
RA Webster R.G.;
RA "Evolution of the H3 influenza virus hemagglutinin from human and
RA nonhuman hosts."
RL J. Virol. 66:1129-1138(1992).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M73771; ; NOT_ANNOTATED_CDS.
CC DR HSSP; P03437; 2VIU.
CC DR InterPro; IPR001364; Hemagglutn.
CC DR Pfam; PF00509; Hemagglutinin; 1.
CC DR PRINTS; PR00329; HEMAGGLUTN12.
CC DR ProDom; PD000225; Hemagglutn; 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; V01087; CAA24271.1; -.
CC DR PDB; 1IBN; 08-AUG-01.
CC DR PDB; 1IBO; 08-AUG-01.
CC DR InterPro; IPR001364; Hemagglutn.
CC DR Pfam; PF00509; Hemagglutinin; 1.
CC DR PRINTS; PR00329; HEMAGGLUTN12.
CC DR ProDom; PD000225; Hemagglutn; 1.
CC -----
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63534 MW; FE19AB6FF9415B89 CRC64;
Query Match 55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAAGFIENGWEGMIDWYG 24
| | | | | | | | | | | | | | | | | | | | | |
Db 346 GLFGAAGFIENGWEGMIDWYG 368
```

```
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63534 MW; FE19AB6FF9415B89 CRC64;
Query Match 55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAAGFIENGWEGMIDWYG 24
| | | | | | | | | | | | | | | | | | | | | |
Db 346 GLFGAAGFIENGWEGMIDWYG 368

RESULT 15
HEMA_IADU3
ID HEMA_IADU3 STANDARD; PRT; 566 AA.
AC P03442;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain]
GN HA.
OS Influenza A virus (strain A/Duck/Ukraine/1/63).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11374;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82025542; PubMed=619439;
RA Fang R., Min Jou W., Huylebroeck D., Devos R., Fiers W.;
RA "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin
RA gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza
RA hemagglutinin."
RL Cell 25:315-323(1981).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; V01087; CAA24271.1; -.
CC DR PDB; 1IBN; 08-AUG-01.
CC DR PDB; 1IBO; 08-AUG-01.
CC DR InterPro; IPR001364; Hemagglutn.
CC DR Pfam; PF00509; Hemagglutinin; 1.
CC DR PRINTS; PR00329; HEMAGGLUTN12.
CC DR ProDom; PD000225; Hemagglutn; 1.
CC -----
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 566 AA; 63530 MW; E70F87F0AE1178F4 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 2.3e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAIAGFIENGWEGMIDGWYG 24
 |||||
 Db 346 GLFGAIAGFIENGWEGMIDGWYG 368
 |||||

Search completed: January 30, 2004, 00:20:46
 Job time : 9.91549 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:52 ; Search time 45.446 Seconds
(without alignments)
249.842 Million cell updates/sec

Title: US-09-461-684C-5
Perfect score: 243
Sequence: 1 CGLFGAIAFIENGWEGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriaph.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	55.1	384	12 Q8JK63	Q8JK63 influenza a
2	134	55.1	566	12 Q98052	Q98052 influenza a
3	134	55.1	566	12 Q8UZ51	Q8UZ51 influenza a
4	134	55.1	566	12 Q8OLN8	Q8OLN8 influenza a
5	134	55.1	566	12 Q67132	Q67132 influenza a
6	134	55.1	566	12 Q67125	Q67125 influenza a
7	134	55.1	566	12 Q8UXR3	Q8UXR3 influenza a
8	134	55.1	566	12 Q91MA7	Q91MA7 influenza a
9	134	55.1	566	12 Q9DHG0	Q9DHG0 influenza a
10	134	55.1	566	12 Q91OM5	Q91OM5 influenza a
11	134	55.1	566	12 Q67126	Q67126 influenza a
12	133	54.7	301	12 Q9DXE3	Q9DXE3 influenza a
13	132	54.3	550	12 Q82499	Q82499 influenza a
14	132	54.3	550	12 Q82498	Q82498 influenza a
15	132	54.3	550	12 Q82753	Q82753 influenza a
16	132	54.3	566	12 Q82496	Q82496 influenza a

17	132	54.3	571	12 Q03909	Q03909 influenza a
18	131	53.9	109	12 Q67050	Q67050 influenza a
19	131	53.9	109	12 Q67053	Q67053 influenza a
20	131	53.9	109	12 Q67051	Q67051 influenza a
21	131	53.9	109	12 Q67052	Q67052 influenza a
22	131	53.9	362	12 Q9QKD3	Q9QKD3 influenza a
23	131	53.9	362	12 Q9QKD1	Q9QKD1 influenza a
24	131	53.9	362	12 Q82513	Q82513 influenza a
25	131	53.9	362	12 Q9QKD2	Q9QKD2 influenza a
26	131	53.9	362	12 Q84174	Q84174 influenza a
27	131	53.9	362	12 Q82517	Q82517 influenza a
28	131	53.9	365	12 Q8DL25	Q8DL25 influenza a
29	131	53.9	367	12 Q9DL22	Q9DL22 influenza a
30	131	53.9	368	12 Q9DL29	Q9DL29 influenza a
31	131	53.9	369	12 Q9DL26	Q9DL26 influenza a
32	131	53.9	369	12 P87689	P87689 influenza a
33	131	53.9	369	12 Q9DL06	Q9DL06 influenza a
34	131	53.9	371	12 Q9DL24	Q9DL24 influenza a
35	131	53.9	371	12 P87685	P87685 influenza a
36	131	53.9	373	12 Q9DL20	Q9DL20 influenza a
37	131	53.9	374	12 Q9DL21	Q9DL21 influenza a
38	131	53.9	375	12 Q9DL27	Q9DL27 influenza a
39	131	53.9	375	12 Q9DL05	Q9DL05 influenza a
40	131	53.9	376	12 Q9DL30	Q9DL30 influenza a
41	131	53.9	376	12 Q9DL04	Q9DL04 influenza a
42	131	53.9	377	12 Q9E7P6	Q9E7P6 influenza a
43	131	53.9	382	12 Q9DL03	Q9DL03 influenza a
44	131	53.9	408	12 Q9E7P5	Q9E7P5 influenza a
45	131	53.9	409	12 Q9Q0L5	Q9Q0L5 influenza a

ALIGNMENTS

RESULT 1

Q8JK63 PRELIMINARY; PRT; 384 AA.
AC Q8JK63; 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hemagglutinin (Fragment).
GN H3HA.
OS Influenza A virus (A/teal/Germany/wv201r/01).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=205472;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/teal/Germany/wv01r/01;
RA Werner O., Starick E., Mueller T., Muehle R.;
RT "Characterisation of avian influenza virus isolates from wild birds from Germany."
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL; AJ506781; CAD4499.1; -!
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 384 384
SQ SEQUENCE 384 AA; 42076 MW; 459731795CA5CE38 CRC64;
Query Match 55.1%; Score 134; DB 12; Length 384;
Best Local Similarity 100.0%; Pred. No. 9.5e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAFIENGWEGMIDGWTG 24

Db 346 GLFGAAGFIENGWEGMIDWYG 368

RESULT 2

Q98052 PRELIMINARY; PRT; 566 AA.
 ID Q98052;
 AC Q98052;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hemagglutinin precursor (Fragment).
 OS Influenzavirus A.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses.
 OX NCBI_TaxID=197911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=81053698; PubMed=6253883;
 RA Both G.W., Sleight M.J.;
 RT "Complete nucleotide sequence of the haemagglutinin gene from a human
 influenza virus of the Hong Kong subtype.";
 RL Nucleic Acids Res. 8:2561-2575(1980).
 RN [2]
 RP SEQUENCE OF 17-344 FROM N.A.
 RX MEDLINE=81194918; PubMed=6164798;
 RA Sleight M.J., Both G.W., Underwood P.A., Bender V.J.;
 RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza
 subtype: Correlation of amino acid changes with alterations in viral
 antigenicity";
 RL J. Virol. 37:845-853(1981).
 RN [3]

RP SEQUENCE OF 17-566 FROM N.A.
 RX MEDLINE=8203276; PubMed=6169843;
 RA Both G.W., Sleight M.J.;
 RT "Conservation and variation in the hemagglutinins of Hong Kong subtype
 influenza viruses during antigenic drift.";
 RL J. Virol. 39:845-853(1981).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; J02135; AAA43189.1; -;
 DR HSSP; P03437; IHGE.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 344 POTENTIAL.
 FT CHAIN 346 566 POTENTIAL.
 SQ SEQUENCE 566 AA; 63414 MW; C447PD465BEAF9 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAAGFIENGWEGMIDWYG 24
 Db 346 GLFGAAGFIENGWEGMIDWYG 368

RESULT 3

Q8U251 PRELIMINARY; PRT; 566 AA.
 ID Q8U251;
 AC Q8U251;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hemagglutinin.
 GN HA.

OS Influenza A virus (A/pet bird/Hong Kong/1559/99(H3N8)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=183665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/pet bird/Hong Kong/1559/99;
 RA Chin P., Shortridge K.F.;
 RT "Characterisation of avian H3 influenza viruses.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AJ427304; CAD20336.1; -;
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 SQ SEQUENCE 566 AA; 63403 MW; F11C91B6A0183484 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAAGFIENGWEGMIDWYG 24
 Db 346 GLFGAAGFIENGWEGMIDWYG 368

RESULT 4

Q8QLN8 PRELIMINARY; PRT; 566 AA.
 ID Q8QLN8;
 AC Q8QLN8;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Haemagglutinin.
 GN HA.
 OS Influenza A virus (A/aquatic bird/Hong Kong/399/99(H3N8)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=183664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/aquatic bird/Hong Kong/399/99;
 RA Chin P., Shortridge K.F.;
 RT "Characterisation of influenza viruses from wild aquatic birds.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AJ427297; CAD20322.1; -;
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 SQ SEQUENCE 566 AA; 63412 MW; 68913C222C97B92E CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLFGAAGFIENGWEGMIDWYG 24
 Db 346 GLFGAAGFIENGWEGMIDWYG 368

RESULT 5

DR	InterPro; IPR001364; Hemagglutn.
DR	Pfam; PF00509; Hemagglutinin; 1.
DR	PRINTS; PR00329; HEMAGGLUTN12.
DR	ProDom; PD000225; Hemagglutin; 1.
KW	Envelope protein; Glycoprotein; Hemagglutinin.
SQ	SEQUENCE 566 AA; 63456 MW; AE556302A9EBB99F CRC64;
 Query Match 55.1%; Score 134; DB 12; Length 566; Best Local Similarity 100.0%; Pred. No. 1.4e-07; Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	2 GLFGAIAGFIENGWEGMIDGWYG 24
DB	346 GLFGAIAGFIENGWEGMIDGWYG 368
 RESULT 7	
Q8UXR3	PRELIMINARY; PRT; 566 AA.
ID	Q8UXR3
AC	Q8UXR3;
DT	01-MAR-2002 (TrEMBLrel. 20, Created)
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE	Hemagglutinin.
GN	HA.
OS	Influenza A virus (A/swine/Potsdam/35/82(H3N2)).
OC	Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
CC	Influenza A viruses; Influenzavirus A.
OX	NCBI_TaxID=183769;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=A/swine/Potsdam/35/82;
RT	Groetzinger I., Sues J., Groetzinger C.;
RA	"Evolution of european human and porcine influenza viruses.";
RL	Submitted (JAN-2000) to the EMBL/Genbank/DDBJ databases.
CC	-1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC	-1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC	-1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC	EMBL; AJ252132; CAC81018.1; --
DR	InterPro; IPR001364; Hemagglutn.
DR	Pfam; PF00509; Hemagglutinin; 1.
DR	PRINTS; PR00329; HEMAGGLUTN12.
DR	ProDom; PD000225; Hemagglutin; 1.
KW	Envelope protein; Glycoprotein; Hemagglutinin.
SQ	SEQUENCE 566 AA; 63529 MW; 6AA4C84B4DD868A CRC64;
 Query Match 55.1%; Score 134; DB 12; Length 566; Best Local Similarity 100.0%; Pred. No. 1.4e-07; Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	2 GLFGAIAGFIENGWEGMIDGWYG 24
DB	346 GLFGAIAGFIENGWEGMIDGWYG 368
 RESULT 8	
Q91MA7	PRELIMINARY; PRT; 566 AA.
ID	Q91MA7
AC	Q91MA7;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE	Hemagglutinin.
OS	Influenza A virus (A/Hong Kong/1/68(H3N2)).
OC	Viruses; sRNA negative-strand viruses; Orthomyxoviridae;
CC	Influenza A viruses; Influenzavirus A.
OX	NCBI_TaxID=108859;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=A/Hong Kong/1/68(H3N2);
RX	MEDLINE=21287244; PubMed=11371620;

RA Brown E.G., Liu H., Kit L.C., Baird S., Nesarallah M.;
 RT "Pattern of mutation in the genome of Influenza A virus on adaptation
 RT to increased virulence in the mouse lung: Identification of functional
 RT themes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AF348176; AAK51718.1; -;
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 DR KW Envelope protein; Glycoprotein; Hemagglutinin.
 SQ SEQUENCE 566 AA; 63387 MW; 01BB0D465BE158E1 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
 |||||
 DB 346 GLFGAAGFIENGWEGMIDGWYG 368

RESULT 9

Q9DHG0
 ID Q9DHG0 PRELIMINARY; PRT; 566 AA.
 AC Q9DHG0;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Haemagglutinin precursor.
 OS Influenza A virus H3N2.
 OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=41857;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=clone 7a;
 RA Mohein M.A., Morris S.J., Smith H., Sweet C.;
 RT "Influenza virus-induced apoptosis: a dual role for viral
 RT neuraminidase.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AJ289703; CAC18525.1; -;
 DR HSP; P03437; 2VIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 DR KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
 FT SIGNAL 1 16 POTENTIAL.
 SQ SEQUENCE 566 AA; 63356 MW; 0BA681929300F72F CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
 |||||
 DB 346 GLFGAAGFIENGWEGMIDGWYG 368

RESULT 10

Q910M5
 ID Q910M5 PRELIMINARY; PRT; 566 AA.

AC Q910M5;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hemagglutinin.
 OS Influenza A virus (A/Hong Kong/1/68 (H3N2)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=108859;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Hong Kong/1/68 (H3N2);
 RX MEDLINE=21287244; PubMed=11371620;
 RA Brown E.G., Liu H., Kit L.C., Baird S., Nesarallah M.;
 RT "Pattern of mutation in the genome of Influenza A virus on adaptation
 RT to increased virulence in the mouse lung: Identification of functional
 RT themes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AF348179; AAK51721.1; -;
 DR EMBL; AF348177; AAK51719.1; -;
 DR EMBL; AF348178; AAK51720.1; -;
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 DR KW Envelope protein; Glycoprotein; Hemagglutinin.
 SQ SEQUENCE 566 AA; 63530 MW; 7CB9F5BAF15E9F4 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
 |||||
 DB 346 GLFGAAGFIENGWEGMIDGWYG 368

RESULT 11

Q67126
 ID Q67126 PRELIMINARY; PRT; 566 AA.
 AC Q67126;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hemagglutinin.
 OS Influenzavirus A.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses.
 OX NCBI_TaxID=197911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Seal/MA/3984/92;
 RX MEDLINE=95146951; PubMed=7844533;
 RA Callan R.J., Early G., Kida H., Hineshaw V.S.;
 RT "The appearance of H3 influenza viruses in seals.";
 RL J. Gen. Virol. 76:199-203(1995).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; L32024; AAA64228.1; -;
 DR HSP; P03437; 2VIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.

KW Envelope protein; Glycoprotein; Hemagglutinin.
 SQ SEQUENCE 566 AA; 63441 MW; 590576CB4CE7D08 CRC64;
 Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
 |||||
 DB 346 GLFGAAGFIENGWEGMIDGWYG 368
 |||||

RESULT 12
 Q9DXE3 PRELIMINARY; PRT; 301 AA.
 AC Q9DXE3;
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
 DE Hemagglutinin (Fragment).
 GN HAI.
 OS Influenza A virus (A/Shorebird/Taiwan/31-4/99).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 OX NCBI_TaxID=140665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Shorebird/Taiwan/31-4/99;
 RA Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee L.H., Shieh H.K.;
 RT "Identification and subtyping of avian influenza virus by reverse
 RT transcription-polymerase chain reaction."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; AF311750; AG33016.1; -;
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON_TER 1 301
 FT NON_TER 301 301
 SQ SEQUENCE 301 AA; 32701 MW; 62A403758B764D57 CRC64;

Query Match 54.7%; Score 133; DB 12; Length 301;
 Best Local Similarity 95.7%; Pred. No. 9.7e-08;
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
 |||||
 DB 250 GLFGAAGFIENGWEGMIDGWYG 272
 |||||

RESULT 13
 Q82499 PRELIMINARY; PRT; 550 AA.
 AC Q82499;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
 DE Hemagglutinins HAI and HA2 (Fragment).
 OS Influenzavirus A.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses.
 OX NCBI_TaxID=197911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Philippines/2/82/BS;
 RA Hartley C.A., Ward A.C., Anders E.M.;
 RT "Virulence of influenza virus for mice is associated with loss of

RT oligosaccharide from the hemagglutinin molecule.";
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; U08859; AAA18782.1; -;
 DR HSSP; P03437; 2VIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON_TER 1 328
 FT NON_TER 330 550
 FT CHAIN 1 328 HA1.
 FT CHAIN 330 550 HA2.
 SQ SEQUENCE 550 AA; 61772 MW; 50BD62B6BE11FD8 CRC64;
 Query Match 54.3%; Score 132; DB 12; Length 550;
 Best Local Similarity 95.7%; Pred. No. 2.3e-07;
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGWEGMIDGWYG 24
 |||||
 DB 330 GLFGAAGFIENGWEGMIDGWYG 352
 |||||

RESULT 14
 Q82498 PRELIMINARY; PRT; 550 AA.
 AC Q82498;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
 DE Hemagglutinins HAI and HA2 (Fragment).
 OS Influenzavirus A.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses.
 OX NCBI_TaxID=197911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Philippines/2/82;
 RA Hartley C.A., Ward A.C., Anders E.M.;
 RT "Virulence of influenza virus for mice is associated with loss of
 RT oligosaccharide from the hemagglutinin molecule.";
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/Philippines/2/82;
 RX MEDLINE=88185444; PubMed=3356226;
 RA Nakajima S., Takeuchi Y., Nakajima K.;
 RT "Location on the evolutionary tree of influenza H3 haemagglutinin
 RT genes of Japanese strains isolated during the 1985-6 season.";
 RL Epidemiol. Infect. 100:301-310 (1988).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; U08858; AAA18781.1; -;
 DR HSSP; P03437; 2VIU.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom; PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON_TER 1 328
 FT NON_TER 330 550
 FT CHAIN 1 328 HA1.
 FT CHAIN 330 550 HA2.
 SQ SEQUENCE 550 AA; 61802 MW; 114413B1CE5A1F6A CRC64;
 Query Match 54.3%; Score 132; DB 12; Length 550;

Best Local Similarity 95.7%; Pred. No. 2.3e-07;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	2	GLFGAIAGFIENGWEGMIDGWYG	24
		:	
Db	330	GIFGAIAGFIENGWEGMIDGWYG	352

RESULT 15

Q82753	PRELIMINARY;	PRT;	550 AA.
IID	Q82753;		
AC	Q82753;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	Haemagglutinin (Fragment).		
DE	Influenza virus.		
OS	Influenza virus.		
OC	Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;		
OC	unclassified Orthomyxoviridae.		
OX	NCBI_TaxID=11309;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=A/Philippines/2/82/BS/ML10;		
RX	MEDLINE=97300854; PubMed=915874;		
RA	Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;		
RA	"Changes in the haemagglutinin molecule of influenza type A		
RA	virus associated with increased virulence for mice."		
RL	Arch. Virol. 142:75-88(1997).		
RL			

RT	"virulence of influenza A virus for mouse lung." ;
RL	Virus Genes 14:187-194(1997).
CC	-I- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC	CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC	-I- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC	(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC	-I- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC	EMBL; U08905; AAC79579.1; -.
DR	HSP; P03437; 2VIU.
DR	InterPro; IPR001364; Hemagglutn.
DR	Pfam; PF00509; Hemagglutinin; 1.
DR	PRINTS; PR00329; HEMAGGLUTN12.
DR	ProDom; PD000225; Hemagglutn; 1.
KW	Envelope protein; Glycoprotein; Hemagglutinin.
FT	NON_TER 1
FT	CHAIN 1 328 HAEMAGGLUTININ HA1.
FT	CHAIN 330 550 HAEMAGGLUTININ HA2.
FT	SEQUENCE 550 AA; 61745 MW; 692AA49DE678AC4BC CRC64;
SQ	

```
Query Match      54.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 2.3e-07;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 2 GLFGAAGFIENGWEGMIDWYG 24
|:||||||| |
Dd 330 GFGAAGFIENGWEGMIDWYG 352

Search completed: January 30, 2004, 00:24:41
Job time : 45.446 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 07:06:28 ; Search time 21 Seconds
(without alignments)
36.636 Million cell updates/sec

Title: SEQ10

Perfect score: 38

Sequence: 1 siinfekl 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76: *
1: Pirl: *
2: Pirl: *
3: Pirl: *
4: Pirl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	386	1	OACH
2	34	89.5	627	2	ovalbumin [validated] - chicken
3	33	86.8	168	2	RNA polymerase I t
4	33	86.8	315	2	hypothetical prote
5	33	86.8	383	2	1-aminocyclopropan
6	32	84.2	280	2	ovalbumin - Japan
7	32	84.2	347	2	conserved hypothet
8	32	84.2	1305	2	cyclin B [imported
9	31	81.6	244	2	hypothetical prote
10	31	81.6	304	2	ABC transport prot
11	31	81.6	307	2	site-specific DNA-
12	31	81.6	309	2	hypothetical prote
13	31	81.6	364	2	hypothetical prote
14	31	81.6	387	2	probable poly(ADP-
15	31	81.6	680	2	hypothetical prote
16	30	78.9	142	2	infected cell prot
17	30	78.9	232	1	50S ribosomal prot
18	30	78.9	347	2	ovalbumin-related
19	30	78.9	349	2	hypothetical prote
20	30	78.9	378	2	probable inner mem
21	30	78.9	440	2	hypothetical prote
22	30	78.9	467	2	hypothetical prote
23	30	78.9	470	2	ATP-dependent DNA
24	30	78.9	481	2	hypothetical prote
25	30	78.9	487	2	conserved hypothet
26	30	78.9	520	2	probable membrane
27	30	78.9	558	2	probable purine-bi
28	30	78.9	610	2	probable membrane
29	30	78.9	842	2	hypothetical prote

30	30	78.9	845	2	D90130
31	30	78.9	1004	2	B69483
32	30	78.9	1115	2	B84476
33	30	78.9	1163	2	G97236
34	30	78.9	2116	2	T49818
35	30	78.9	2469	2	H36812
36	30	78.9	4092	1	S38128
37	29	76.3	69	2	A96696
38	29	76.3	88	4	S54022
39	29	76.3	110	2	F97094
40	29	76.3	141	2	H69044
41	29	76.3	143	2	S45537
42	29	76.3	170	2	G86356
43	29	76.3	209	2	D86882
44	29	76.3	237	2	F64506
45	29	76.3	275	1	C69030

ALIGNMENTS

RESULT 1

OACH

ovalbumin [validated] - chicken
C:Species: Gallus gallus (Chicken)

C:Date: 31-Dec-1979 #sequence revision 30-Jun-1993 #text change 15-Sep-2000
C:Accession: A90455; I50402; I50605; A93197; A93827; A90093; A30092; A61297; A42793; A01

R:Woo, S.L.C.; Beattie, W.G.; Catterall, J.F.; Dugaiczky, A.; Staden, R.; Brownlee, G.G.
Biochemistry 20, 6437-6446, 1981

A:Title: Complete nucleotide sequence of the chicken chromosomal ovalbumin gene and its

A:Reference number: A90455; MUID:82069038; PMID:6272839

A:Accession: A90455

A:Molecule type: DNA

A:Residues: 1-386 <WOO>

A:Cross-references: EMBL:V00438; NID:G63719; PIDN:CRAA23716.1; PID:9808974

A:Note: a number of silent polymorphic sites are identified and discussed

A:Note: Thr-188 is also predicted

R:Catterall, J.F.; O'Malley, B.W.; Robertson, M.A.; Staden, R.; Tanaka, Y.; Brownlee, G.

Nature 275, 510-513, 1978

A:Title: Nucleotide sequence homology at 12 intron-exon junctions in the chick ovalbumin

A:Reference number: I50402; MUID:79010682; PMID:692731

A:Accession: I50402

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-386 <CAT>

A:Cross-references: GB:M34352; NID:g212501; PIDN:AAA48998.1; PID:g212503

R:Robertson, M.A.; Staden, R.; Tanaka, Y.; Catterall, J.F.; O'Malley, B.W.; Brownlee, G.

Nature 278, 370-372, 1979

A:Title: Sequence of three introns in the chick ovalbumin gene.

A:Reference number: I50605; MUID:79135070; PMID:423993

A:Accession: I50605

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4, 'A', 6-118, 'P', 120-155 <ROB>

A:Cross-references: EMBL:V00382; NID:G63051; PIDN:CRAA23681.1; PID:G63052

R:McReynolds, L.; O'Malley, B.W.; Nisbet, A.D.; Fothergill, J.E.; Givol, D.; Fields, S.;

Nature 273, 723-728, 1978

A:Title: Sequence of chicken ovalbumin mRNA.

A:Reference number: A93197; MUID:78199842; PMID:661981

A:Accession: A93197

A:Molecule type: mRNA

A:Residues: 1-386 <MCR>

A:Cross-references: EMBL:V00383; NID:G63053

A:Note: a minor component has Asp-312

R:Palmiter, R.D.; Gagnon, J.; Walsh, K.A.

Proc. Natl. Acad. Sci. U.S.A. 75, 94-98, 1978

A:Title: Ovalbumin: a secreted protein without a transient hydrophobic leader sequence.

A:Reference number: A93827; MUID:78116057; PMID:272676

A:Accession: A93827

A:Molecule type: protein

A:Residues: 2-33, 'X', 35-36 <PAL>

R:Thompson, E.O.P.; Fisher, W.K.

Aust. J. Biol. Sci. 31, 443-446, 1978

A:Title: A correction and extension of the acetylated amino terminal sequence of ovalbumin
A:Reference number: A90093; PMID:79186958; PMID:751625
A:Accession: A90093
A:Molecule type: protein
A:Residues: 2-17 <TH1>
R:Thompson, E.O.P.; Fisher, W.K.
Aug. J. Biol. Sci. 31, 433-442, 1978
A:Title: Amino acid sequences containing half-cysteine residues in ovalbumin.
A:Reference number: A90092; PMID:79186957; PMID:751624
A:Accession: A90092
A:Molecule type: protein
A:Residues: 6-17;30-36;61-79;116-124;367-374;380-386 <TH2>
R:Tsunasawa, S.; Narita, K.
J. Biochem. 92, 607-613, 1982
A:Title: Micro-identification of amino-terminal acetylamino acids in proteins.
A:Reference number: A61297; PMID:83056735; PMID:6754709
A:Accession: A61297
A:Molecule type: protein
A:Residues: 2-6 <TGU>
R:Takahashi, N.; Hirose, M.
J. Biol. Chem. 267, 11565-11572, 1992
A:Title: Reversible denaturation of disulfide-reduced ovalbumin and its reoxidation gene
A:Reference number: A42793; PMID:92283876; PMID:1597484
A:Accession: A42793
A:Molecule type: protein
A:Residues: 60-73, 'X', 75-85;112-119, 'EX', 122-123 <TAX>
R:Stein, P.E.; Leslie, A.G.W.
Submitted to the Brookhaven Protein Data Bank, November 1990
A:Reference number: A50294; PDB:1OYA
A:Contents: annotation; X-ray crystallography, 1.95 angstroms, residues 2-386
R:Stein, P.E.; Leslie, A.G.W.; Finch, J.T.; Carrell, R.W.
J. Mol. Biol. 221, 941-959, 1991
A:Title: Crystal structure of uncleaved ovalbumin at 1.95 Angstroms resolution.
A:Reference number: A58761; PMID:92046044; PMID:1942038
A:Contents: annotation; X-ray crystallography, 1.95 angstroms
C:Genetics:
A:Introns: 56/3; 73/3; 116/3; 156/1; 203/3; 255/3
C:Superfamily: antithrombin III
C:Keywords: acetylated amino end; glycoprotein; phosphoprotein
F:2-384/Product: ovalbumin #status experimental <MAR>
F:2/Modified site: acetylated amino end (Gly) (in mature form) #status experimental
F:69.345/Binding site: phosphate (Ser) (covalent) #status experimental
F:74-121/Disulfide bonds: #status experimental
F:293/Binding site: carbohydrate (Asn) (covalent) #status experimental
Query Match 100.0%; Score 38; DB 1; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 258 SIINFEKL 265
RESULT 2
S37954
RNA polymerase I transcription factor RRN3 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKL125W
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 13-Mar-1998
R:Ramezani Rad, M.; Xu, G.; Kirchrath, L.; Fritz, C.; Keuchel, H.; Hollenberg, C.P.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37953
A:Accession: S37954
A:Molecule type: DNA
A:Residues: 1-627 <RAM>
A:Cross-references: EMBL:Z28125; NID:9486210; PID:9486211; MIPS:YKL125W
A:Experimental source: strain S288C
R:Yamamoto, R.T.; Nogi, Y.; Dodd, J.A.; Nomura, M.
EMBO J. 15, 3964-3973, 1996
A:Title: RRN3 gene of Saccharomyces cerevisiae encodes an essential RNA polymerase I tra
A:Reference number: S71600; PMID:96324404; PMID:8670901

A:Accession: S71600
A:Molecule type: DNA
A:Residues: 1-627 <YAM>
C:Genetics:
A:Gene: SGD:RRN3
A:Cross-references: SGD:S0001608; MIPS:YKL125W
A:Map position: 11L
C:Keywords: nucleus
Query Match 89.5%; Score 34; DB 2; Length 627;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEK 7
DB 118 SIINFEK 124
RESULT 3
P90095
Hypothetical protein orf168 [imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: P90095
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re
Nature 410, 1091-1096, 2001
A:Title: The highly reduced genome of an enslaved algal nucleus.
A:Reference number: A99082; PMID:11323671; PMID:11323671
A:Accession: P90095
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-168 <DOU>
A:Cross-references: GB:AF165818; NID:913794518; PIDN:AAK39893.1; GSPDB:GN00150
C:Genetics:
A:Gene: orf168
A:Map position: 1
A:Genome: nucleomorph
C:Keywords: nucleomorph
Query Match 86.8%; Score 33; DB 2; Length 168;
Best Local Similarity 75.0%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 108 SIINFEKI 115
RESULT 4
TI0818
1-aminocyclopropane-1-carboxylate oxidase (EC 1.4.3.-) - kidney bean
C:Species: Phaseolus vulgaris (kidney bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 19-May-2000
C:Accession: TI0818
R:Pidgen, C.M.; Facchini, P.J.; Reid, D.M.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z17172
A:Accession: TI0818
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-315 <PID>
A:Cross-references: EMBL:AF053354; NID:93037046; PIDN:AAC12934.1; PID:93037047
A:Experimental source: cultivar Taylor bush bean; leaf
C:Genetics:
A:Gene: ACO1
C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase
C:Keywords: ethylene biosynthesis; iron; metalloprotein; oxidoreductase
F:319,177,234/Binding site: iron (His) #status predicted
Query Match 86.8%; Score 33; DB 2; Length 315;
Best Local Similarity 85.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
QY      2 IINFEKL 8
      :|||||
Db      6 VINFEKL 12

RESULT 5
S11433
ovalbumin - Japanese quail
C/Species: Coturnix coturnix japonica (Japanese quail)
C/Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C/Accession: S11433
R/Mucha, J.; Klaudiny, J.; Klaudinyova, V.; Hanes, J.; Simuth, J.
Nucleic Acids Res. 18, 5553, 1990
A/Title: The sequence of Japanese quail ovalbumin cDNA.
A/Reference number: S11433; MUID:91016850; PMID:2216734
A/Accession: S11433
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-383 <MUC>
A/Cross-references: EMBL:X53964; NID:g62643; PIDN:CAA37916.1; PID:g62644
C/Superfamily: antithrombin III

Query Match      86.8%; Score 33; DB 2; Length 383;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIINFEKL 8
      :|||||
Db     258 SIISFEKL 265

RESULT 6
G87349
conserved hypothetical protein CC0810 [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: G87349
R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: G87349
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-260 <STO>
A/Cross-references: GB:AB005673; NID:g13422057; PIDN:AAK22795.1; GSPDB:GN00148
C/Genetics:
A/Gene: CC0810

Query Match      84.2%; Score 32; DB 2; Length 260;
Best Local Similarity 87.5%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 SIINFEKL 8
      :|||||
Db     215 SIINFEAL 222

RESULT 7
A99989
Cyclin B [imported] - Guillardia theta nucleomorph
C/Species: nucleomorph Guillardia theta
A/Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C/Accession: A99989
R/Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei
Nature 410, 1091-1096, 2001
A/Title: The highly reduced genome of an enslaved algal nucleus.
A/Reference number: A99082; MUID:11323671; PMID:11323671
A/Accession: A99989
```

```
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-347 <DOU>
A/Cross-references: GB:AF165818; NID:g13794469; PIDN:AAK39844.1; GSPDB:GN00150
C/Genetics:
A/Gene: CYCB
A/Map position: 1
A/Genome: nucleomorph
C/Keywords: nucleomorph

Query Match      84.2%; Score 32; DB 2; Length 347;
Best Local Similarity 62.5%; Pred. No. 40;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIINFEKL 8
      :|||||
Db      86 NVLNFEKL 93

RESULT 8
T23314
hypothetical protein T14G10.2 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C/Accession: T23314; T24919
R/Wild, A.
submitted to the EMBL Data Library, February 1996
A/Reference number: Z19725
A/Accession: T23314
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1305 <WIL>
A/Cross-references: EMBL:Z69664; PIDN:CAA93519.1; GSPDB:GN00022; CESP:T14G10.2
A/Experimental source: clone K04D7
R/Wild, A.
submitted to the EMBL Data Library, January 1996
A/Reference number: Z19954
A/Accession: T24919
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1305 <WIL>
A/Cross-references: EMBL:Z68880; PIDN:CAA93100.1; GSPDB:GN00022; CESP:T14G10.2
A/Experimental source: clone T14G10
C/Genetics:
A/Gene: CESP:T14G10.2
A/Map position: 4
A/Introns: 450/1; 463/2; 696/2; 763/2; 843/2; 935/3; 1012/1; 1091/1; 1143/1; 1189/2; 12;

Query Match      84.2%; Score 32; DB 2; Length 1305;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 IINFEKL 8
      :|||||
Db     903 LINFEKL 909

RESULT 9
T06961
ABC transport protein homolog - Cyanophora paradoxa cyanelle
C/Species: cyanelle Cyanophora paradoxa
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 17-Mar-2000
C/Accession: T06961
R/Stirewalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A/Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A/Reference number: Z15840
A/Accession: T06961
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-244 <STI>
A/Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81304.1; PID:g1016217
A/Experimental source: strain Pringsheim 18555
```

C:Genetics:
A:Genome: cyanelle
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: cyanelle

Query Match 81.6%; Score 31; DB 2; Length 244;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
|||||:
DB 204 SIINFEDEL 211

RESULT 10
C64109
site-specific DNA-methyltransferase (cytosine-specific) (EC 2.1.1.73) - Haemophilus infl
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 21-Jul-2000
C:Accession: C64109
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: C64109
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-304 <TIGR>
A:Cross-references: GB:U32784; GB:I42023; NID:93212210; PIDN:AAC22700.1; PID:gl574073; T
C:Superfamily: site-specific methyltransferase (cytosine-specific) EcoRII
C:Keywords: methyltransferase; restriction modification system; S-adenosylmethionine

Query Match 81.6%; Score 31; DB 2; Length 304;
Best Local Similarity 85.7%; Pred. No. 56;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEK 7
|||||:
DB 291 AIINFEK 297

RESULT 11
E71206
hypothetical protein PH1919 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: E71206
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: E71206
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-307 <RAW>
A:Cross-references: GB:AP000007; NID:93236134; PIDN:BA31044.1; PID:G3258361
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1919

Query Match 81.6%; Score 31; DB 2; Length 307;
Best Local Similarity 75.0%; Pred. No. 57;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
|||||:
DB 3 NIINFEEL 10

RESULT 12

H90548
hypothetical protein MYPV 2960 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: H90548
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: H90548
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-309 <KUR>
A:Cross-references: GB:AL445566; PID:gl4089710; PIDN:CAC13469.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPV 2960
A:Genetic code: SGC3

Query Match 81.6%; Score 31; DB 2; Length 309;
Best Local Similarity 71.4%; Pred. No. 57;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEK 7
|||||:
DB 287 SVLNFEK 293

RESULT 13

A84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84726
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, I.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84726
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <STO>
A:Cross-references: GB:AB002093; NID:94887750; PIDN:AAD32286.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g31860
A:Map position: 2

Query Match 81.6%; Score 31; DB 2; Length 364;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IINFEKL 8
|||||:
DB 88 IINFDKL 94

RESULT 14

T34441
hypothetical protein K11H12.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34441
R:Bradshaw, H.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid K11H12.
A:Reference number: Z21526
A:Accession: T34441
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A;Residues: 1-397 <BRA>
A;Cross-references: EMBL:U88168; PIDN:AAC24400.1; GSPDB:GN00022; CESP:K1IH12.3
A;Experimental source: strain Bristol N2; clone K1IH12
C;Genetics:
A;Gene: CESP:K1IH12.3
A;Map position: 4
A;Introns: 39/3; 68/2; 118/1; 206/2; 280/3

Query Match 81.6%; Score 31; DB 2; Length 397;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
| | | | |
Db 190 SIINFSKI 197

RESULT 15
T42923
infected cell protein - ateline herpesvirus 3 (strain 73)
C;Species: ateline herpesvirus 3
A;Variety: strain 73
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 05-May-2000
C;Accession: T42923
R;Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A;Description: Primary structure of the herpesvirus ateles genome.
A;Reference number: 222274
A;Accession: T42923
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-680 <ALB>
A;Cross-references: EMBL:AF083424; PIDN:AAC95539.1
A;Experimental source: strain 73
C;Genetics:
A;Note: orf07
C;Superfamily: herpesvirus infected cell protein ICP18.5

Query Match 81.6%; Score 31; DB 2; Length 680;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEX 7
| : | | | |
Db 618 SVINFEX 624

Search completed: January 30, 2004, 07:09:46
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 07:06:27 ; Search time 12 Seconds

(without alignments)
31.351 Million cell updates/sec

Title: SEQ10

Perfect score: 38

Sequence: 1 siinfekl 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	385	1	OVAL CHICK
2	34	89.5	627	1	RN3 YEAST
3	33	86.8	382	1	OVAL COTJA
4	32	84.2	267	1	RM09 HUMAN
5	31	81.6	242	1	TP15 MYCFL
6	31	81.6	244	1	YKXD CRAPA
7	31	81.6	304	1	MTH5 HABIN
8	30	78.9	231	1	RL1 BUCAP
9	30	78.9	232	1	OVAX CHICK
10	30	78.9	349	1	Y567 HELPJ
11	30	78.9	354	1	SERC DROME
12	30	78.9	527	1	FPCK FUSNN
13	30	78.9	558	1	MNT2 YEAST
14	30	78.9	842	1	DP3A THEMA
15	30	78.9	1163	1	SBCC CLOAB
16	30	78.9	2469	1	TEGU HSVSA
17	30	78.9	4092	1	DYHC YEAST
18	29	76.3	143	1	PP1B BACSU
19	29	76.3	237	1	YG56 METJA
20	29	76.3	394	1	C91 YEAST
21	29	76.3	445	1	RPN5 YEAST
22	29	76.3	453	1	EX7L RICPR
23	29	76.3	482	1	V233 FOWPV
24	29	76.3	489	1	QJ503 fowlpox vir
25	29	76.3	749	1	PCRA LEUCI
26	29	76.3	1002	1	HPS3 MOUSE
27	29	76.3	1233	1	SMC1 SCHPO
28	29	76.3	1234	1	YXK5 CAEEL
29	29	76.3	2054	1	YCP2 PINTH
30	28	73.7	102	1	CYTI ORISA
31	28	73.7	124	1	PA25 AGKHP
32	28	73.7	177	1	YA50 METJA
33	28	73.7	214	1	PYRE_PASMU

ALIGNMENTS

RESULT 1
OVAL_CHICK STANDARD; PRT; 385 AA.
AC P01012; 222 1 BID2 YERPE
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ovalbumin (Plakalbumin) (Allergen Gal d 2) (Gal d II).
OS Gallus Gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82069038; PubMed=6272839;
RA Woo S.L.C., Beattie W.G., Catterall J.F., Dugaiczky A., Staden R.,
RA Brownlee G.G., O'Malley B.W.;
RT "Complete nucleotide sequence of the chicken chromosomal ovalbumin
gene and its biological significance."
RL Biochemistry 20:6437-6446(1981).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=78199842; PubMed=661981;
RA McReynolds L., O'Malley B.W., Nisbet A.D., Fothergill J.E., Givol D.,
RA Fields S., Robertson M., Brownlee G.G.;
RT "Sequence of chicken ovalbumin mRNA."
RL Nature 273:723-728 (1978).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=79010682; PubMed=692731;
RA Catterall J.F., O'Malley B.W., Robertson M.A., Staden R.,
RA Tanaka Y., Brownlee G.G.;
RT "Nucleotide sequence homology at 12 intron-exon junctions in the
chick ovalbumin gene."
RL Nature 275:510-513 (1978).
RN [4]
RP SEQUENCE OF 1-35.
RX MEDLINE=78116057; PubMed=272676;
RA Palminter R.D., Gagnon J., Walsh K.A.;
RT "Ovalbumin: a secreted protein without a transient hydrophobic leader
sequence."
RL Proc. Natl. Acad. Sci. U.S.A. 75:94-98 (1978).
RN [5]
RP SEQUENCE OF 1-16.
RX MEDLINE=79186958; PubMed=751625;
RA Thompson E.O.P., Fisher W.K.;
RT "A correction and extension of the acetylated amino terminal sequence
of ovalbumin."
RL Aust. J. Biol. Sci. 31:443-446 (1978).
RN [6]
RP SEQUENCE OF 5-16; 29-35; 60-78; 115-123; 366-373 AND 379-385.
RX MEDLINE=79186957; PubMed=751624;
RA Thompson E.O.P., Fisher W.K.;
RT "Amino acid sequences containing half-cysteine residues in ovalbumin."
RL Aust. J. Biol. Sci. 31:433-442 (1978).

34 28 73.7 222 1 BID2 YERPE
35 28 73.7 240 1 YDSD_SCHPO
36 28 73.7 242 1 TPIS_MYCHY
37 28 73.7 269 1 H1S9_LACLA
38 28 73.7 278 1 TNF6 RAT
39 28 73.7 279 1 TNF6_MOUSE
40 28 73.7 306 1 COA4_STRP3
41 28 73.7 306 1 COA4_STRP8
42 28 73.7 306 1 COA4_STRP3
43 28 73.7 309 1 LDH2_LACPL
44 28 73.7 315 1 ACC1_LYCBS
45 28 73.7 316 1 ACC2_LYCBS
O9agd4 versinia pe
Q10446 schizosacch
P50920 mycoplasma
Q02150 lactococcus
P36940 rattus norv
P41047 mus musculu
O8k7c7 streptococc
Q8pov9 streptococc
Q992h1 streptococc
P59390 lactobacill
P05116 lycopersico
P07920 lycopersico


```

OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Oviduct;
RX MEDLINE=91016850; PubMed=2216734;
RA Mucha J., Klaudiny J., Klaudinyova V., Hanes J., Simuth J.;
RT "The sequence of Japanese quail ovalbumin cDNA.";
RL Nucleic Acids Res. 18:5553-5553(1990).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53964; CAA37916.1; -
DR PIR; S11433; S11433.
DR HSSP; P01012; IOVA.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR SMART; SM00093; SERPIN; 1.
DR PROSITE; PS00284; SERPIN; 1.
KW Serpin; Glycoprotein.
FT INIT MET 0
FT DISULFID 73 120 BY SIMILARITY.
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 311 311 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT ACT SITE 352 353 REACTIVE BOND HOMOLOG.
SQ SEQUENCE 382 AA; 42108 MW; A4BB59A1BAEF316 CRC64;

Query Match 86.8%; Score 33; DB 1; Length 382;
Best Local Similarity 87.5%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFELK 8
Db 257 SIISFEKL 264

RESULT 4
RM09 HUMAN
ID RM09 HUMAN STANDARD; PRT; 267 AA.
AC Q9BYD2; Q9BSW8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 60S ribosomal protein L9, mitochondrial precursor (L9mt).
GN MRPL9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21293042; PubMed=11279069;
RA Suzuki T., Terasaki M., Takemoto-Hori C., Hanada T., Ueda T., Wada A.,
RA Watanabe K.;
RT "Structural compensation for the deficit of rRNA with proteins in the
RT mammalian mitochondrial ribosome. Systematic analysis of protein
RT components of the large ribosomal subunit from mammalian
RT mitochondria.";
RL J. Biol. Chem. 276:21724-21736(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;

```

```

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Schenken C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toehiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M.,
RA Butterfield A., Schain J.E., Jones S.J.M., Marra M.A.,
RA Schnurch A., Krzywinski M.I., Skalska U.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: BELONGS TO THE L9P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB049636; BAB40841.1; -;
CC EMBL; BC004517; AAH04517.1; -;
CC Genew; HGNC:14277; MRPL9.
CC InterPro; IPR000244; Ribosomal L9.
CC Pfam; PF01281; Ribosomal L9 N.
KW Ribosomal protein; Mitochondrion; Trans peptide.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 267 608 RIBOSOMAL PROTEIN L9.
FT CONFLICT 210 210 A -> E (IN REF. 2).
SQ SEQUENCE 267 AA; 30185 MW; 346C254220FFD1B4 CRC64;

Query Match 84.2%; Score 32; DB 1; Length 267;
Best Local Similarity 71.4%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEK 7
Db 236 SVNFEK 242

RESULT 5
TPIS MYCFL STANDARD; PRT; 242 AA.
ID TPIS MYCFL STANDARD; PRT; 242 AA.
AC P48779;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Triosephosphate isomerase (EC 5.3.1.1) (TIM).
GN TPIA OR TPI.
OS Mycoplasma flocculare.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2128;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27399;
RA Xiang H., McIntosh M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBSJ databases.
CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = Glycerone
CC phosphate.
CC -1- PATHWAY: Plays an important role in several metabolic pathways.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U20509; AAA62167.1; -;
CC HSP; P00943; 2BTM.
CC HAMAP; MF 00147; -; 1.
CC InterPro; IPR000652; Triophos_ismrse.
CC Pfam; PF00121; TIM; 1.
CC ProDom; PD001005; Triophos_ismrse; 1.
CC TIGRFAMs; TIGR00419; tim; 1.
CC PROSITE; PS00171; TIM; 1.
KW Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW Pentose shunt.
FT ACT_SITE 98 98 BY SIMILARITY.
FT ACT_SITE 167 167 BY SIMILARITY.
SQ SEQUENCE 242 AA; 26969 MW; E1E8560E2DA18FA1 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 242;
Best Local Similarity 71.4%; Pred. No. 18;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEK 7
Db 155 SVLNFEX 161

RESULT 6
YCXD CYAPA STANDARD; PRT; 244 AA.
ID YCXD CYAPA STANDARD; PRT; 244 AA.
AC P48334;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable ABC transporter ATP-binding protein in ycf23-apoF intergenic
DE region (ORF244).
OS Cyanophora paradoxa.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluter W.M., Chung S., Neumann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluter W.M., Chung S., Neumann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluter W.M., Chung S., Neumann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27399;
RA Xiang H., McIntosh M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBSJ databases.
CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = Glycerone
CC phosphate.
CC -1- PATHWAY: Plays an important role in several metabolic pathways.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.

```

```

CC -----
DR EMBL; U30821; AAA81304.1; --
DR FR; T06961; T06961.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA_1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein; ATP-binding; Transport; Cyanelle.
FT NP BIND 41 48 ATP (POTENTIAL).
SQ SEQUENCE 244 AA; 27747 MW; 4C5B357FF9C55D3B CRC64;

Query Match 81.6%; Score 31; DB 1; Length 244;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
DB 204 SIINFDEL 211

RESULT 7
MTHS_HABIN STANDARD; PRT; 304 AA.
AC P45030;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Modification methylase Hindv (EC 2.1.1.73) (Cytosine-specific
DE methyltransferase Hindv) (M.Hindv).
DE HINDVM OR H11041.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA whole-genome random sequencing and assembly of Haemophilus influenzae
RA Rd.;
RL Science 269:496-512(1995).
CC -1- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC GRCGVC. CAUSES SPECIFIC METHYLATION ON C-? ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE HINDV ENDONUCLEASE.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA cytosine = S-
CC adenosyl-L-homocysteine + DNA 5-methylcytosine.
CC -1- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32784; AAC22700.1; --
DR FR; C64109; C64109.
DR HSSP; O14717; 1G55.
DR REBASE; 3574; M.Hindv.
DR TIGR; H11041; --

```

```

DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF00145; DNA_methylase; 1.
DR PRINTS; PR00105; C5METTRFRASE.
DR TIGRFAMs; TIGR00675; dcm; 1.
DR PROSITE; PS00094; C5_MTASE_1; 1.
DR PROSITE; PS00095; C5_MTASE_2; 1.
KW Hypothetical protein; Transferase; Methyltransferase;
KW Restriction system; Complete proteome.
FT ACT SITE 75 75
SQ SEQUENCE 304 AA; 34365 MW; 03DA1EAB27C84BBD CRC64;

Query Match 81.6%; Score 31; DB 1; Length 304;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEK 7
DB 291 AIINFEK 297

RESULT 8
RL1_BUCAP STANDARD; PRT; 231 AA.
AC Q8K367;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L1.
DE RPLA OR BUSG038.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379(2002).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
CC LOCATED IN THE NEIGHBORHOOD OF THE SITE WHERE ELONGATION FACTOR TU
CC IS BOUND TO THE RIBOSOME (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L1P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB014080; AA067609.1; --
DR InterPro; IPR005878; L1_bact chl.
DR InterPro; IPR002143; Ribosomal_L1.
DR Pfam; PF00687; Ribosomal_L1; 1.
DR ProDom; PD001314; Ribosomal_L1; 1.
DR TIGRFAMs; TIGR01169; rplA_bact; 1.
DR PROSITE; PS01199; RIBOSOMAL_L1; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 231 AA; 25609 MW; 4D234DACF932A3C2 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 231;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 INFEKL 8
DB 15 INFEKL 20

RESULT 9

```

```

OVAX_CHICK
ID OVAX_CHICK STANDARD; PRT; 232 AA.
AC P01013;
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Gene X protein (Ovalbumin-related) (Fragment).
GN X.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81022623; PubMed=7418002;
RA Heilig R., Perrin F., Gannon F., Mandel J.L., Chambon P.;
RT "The ovalbumin gene family: structure of the X gene and evolution of
RL Cell 20:625-637(1980).
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J00920; AAA68881.1; -.
CC EMBL: J00918; AAA68881.1; JOINED.
CC EMBL: J00919; AAA68881.1; JOINED.
CC EMBL: V00385; CAA23683.1; -.
CC EMBL: V00386; CAA23684.1; -.
CC EMBL: V00387; CAA23685.1; -.
CC PIR: A01243; DXCH.
CC HSP: P01012; IOVA.
CC InterPro: IPR000215; Serpin.
CC Pfam: PF00079; serpin; 1.
CC SMART: SM00093; SERPIN; 1.
CC PROSITE: PS00284; SERPIN; 1.
KW Serpin.
FT NON_TER 1 1
SQ SEQUENCE 232 AA; 26291 MW; 6B5B86EC4D3B9195 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 232;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 INFEKL 8
DB 104 INFEKL 109
|||||

RESULT 10
Y567_HELPFU
ID Y567_HELPFU STANDARD; PRT; 349 AA.
AC Q9ZLR4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein JHP0514.
GN JHP0514.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,

```

```

RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RL Nature 387:176-180(1999)
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPP0118 (PERM) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: A0001484; AAD06090.1; -.
CC PIR: H71923; H71923.
CC InterPro: IPR002549; UPP0118.
CC Pfam: PF01594; UPP0118; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 6 26
FT TRANSMEM 27 47
FT TRANSMEM 56 76
FT TRANSMEM 143 163
FT TRANSMEM 195 215
FT TRANSMEM 224 244
FT TRANSMEM 258 278
FT TRANSMEM 300 320
SQ SEQUENCE 349 AA; 39804 MW; A1846D48CB3A8B6F CRC64;

Query Match 78.9%; Score 30; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 INFEKL 8
DB 85 INFEKL 90
|||||

RESULT 11
SERC_DROME
ID SERC_DROME STANDARD; PRT; 364 AA.
AC Q9VAN0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable phosphoserine aminotransferase (EC 2.6.1.52) (PSAT).
GN ESTS:39C10S OR CG11899.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer G., Champe M., Pfeiffer B.D.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balwle R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```



```

RP CHARACTERIZATION.
RA MEDLINE=99453866; PubMed=10521541;
RA Romero P.A., Lussier M., Veronneau S., Sdicu A.M., Herscovics A.,
RA Bussey H.;
RT "Mnt2p and Mnt3p of Saccharomyces cerevisiae are members of the Mntlp
RT family of alpha-1,3-mannosyltransferases responsible for adding the
RT terminal mannose residues of O-linked oligosaccharides.";
RL Glycobiology 9:1045-1051(1999).
CC -!- FUNCTION: Mannosyltransferase involved in adding the 4th and 5th
CC mannose residues of O-linked glycans.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- SIMILARITY: BELONGS TO THE MNT1/MNT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X94357; CRA64130.1; -.
DR EMBL; Z72779; CRA96977.1; -.
DR PIR; S61604; S61604.
DR SGD; S0003226; MNT2.
DR GO; GO:0000033; F:alpha-1,3-mannosyltransferase activity; IDA.
DR GO; GO:0006493; P:O-linked glycosylation; IDA.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack.
FT DOMAIN 1 6
FT TRANSMEM 7 27
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT LUMENAL (POTENTIAL).
FT DOMAIN 28 558
FT CARBOHYD 187 187
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 558 AA; 64852 MW; 3E58ED62B4E291B6 CRC64;
Query Match 78.9%; Score 30; DB 1; Length 558;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 INFEKL 8
DB 109 INFEKL 114
RESULT 14
DP3A THEME STANDARD; PRT; 842 AA.
AC Q9ZHG4.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase III alpha subunit (EC 2.7.7.7).
GN DNAE OR TM0461.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=23336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99045593; PubMed=9826752;
RA Huang Y.P., Ito J.;
RT "The hyperthermophilic bacterium Thermotoga maritima has two different
RT classes of family C DNA polymerases: evolutionary implications.";
RL Nucleic Acids Res. 26:5300-5309(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,

```

```

RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
CC epsilon and theta chains) that associates with a tau subunit. This
CC core dimerizes to form the POLII' complex. POLII' associates
CC with the gamma complex (composed of gamma, delta, delta', psi and
CC chi chains) and with the beta chain to form the complete DNA
CC polymerase III complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-C FAMILY. DNAE
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF063188; AAC80434.1; -.
DR EMBL; AE001724; AAD35546.1; -.
DR PIR; E72373; E72373.
DR TIGR; TM0461; -.
DR InterPro; IPR003141; PHP_N.
DR SMART; IPR004805; PolC_alpha.
DR SMART; SM00481; POLIITAC; 1.
DR TIGRFRAMS; TIGR00594; PolC; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW Complete proteome.
FT CONFLICT 377 377 I -> M (IN REF. 1).
SQ SEQUENCE 842 AA; 96499 MW; 19BB813C085C277 CRC64;
Query Match 78.9%; Score 30; DB 1; Length 842;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 14 SVVRFELK 21
RESULT 15
SBCC CLOAB STANDARD; PRT; 1163 AA.
ID SBCC CLOAB
AC Q97FKL;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nuclease sbcCD subunit C.
GN SBCC OR CAC2736.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing

```

```

RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838 (2001)).
CC -!- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
CC can inhibit DNA replication and are intermediates in certain DNA
CC recombination reactions. The complex acts as a 3'->5' double
CC strand exonuclease that can open hairpins. It also has a 5'
CC single-strand endonuclease activity (By similarity).
CC -!- SUBUNIT: Heterodimer of sbcc and sbcd (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL; AE007771; AAK80682.1; -.
DR PIR; G97236; G97236.
DR InterPro; IPR003439; ABC_transporter.
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; DNA replication;
KW DNA recombination; ATP-binding; Coiled coil; Complete proteome.
FT NP BIND 35 42 ATP (POTENTIAL).
FT DOMAIN 197 415 COILED COIL (POTENTIAL).
FT DOMAIN 446 1003 COILED COIL (POTENTIAL).
SQ SEQUENCE 1163 AA; 135507 MW; C85F0BD2215D7A92 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 1163;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 INFEKL 8
Db 20 INFEKL 25

```

Search completed: January 30, 2004, 07:07:49
 Job time : 14 secs

! FINDPATTERNS on geneseq:* allowing 0 mismatches

```

1 1 C(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V)
AAB13781 ck: 1303 len: 25 1 Aab13781 Soluble peptide antigen pEA. 11/20
C(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I
C(E) (A) {5} (E) (A) {5} (E) (A) {5}
1: CEAAAAAAAAAAAAA EAAAA
AAB13783 ck: 4553 len: 45 1 Aab13783 Soluble tandem pEA/ pX peptide con
C(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I
C(E) (A) {5} (E) (A) {5} (E) (A) {5}
1: CEAAAAAAAAAAAAA EAAAA

```

Databases searched:

Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds: 2

Total length: 158,726,570

Total sequences: 1,107,863

CPU time: 05:59.54

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 07:15:53 ; Search time 71 Seconds
(without alignments)
17.885 Million cell updates/sec

Title: SEQ10
Perfect score: 38
Sequence: 1 siinfekl 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 125

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 250 summaries

Database : A_Geneseq_13Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	8	15	AA19857996
2	38	100.0	8	16	AA1983938
3	38	100.0	8	17	AA1989157
4	38	100.0	8	18	AA199555
5	38	100.0	8	18	AA194087
6	38	100.0	8	18	AA194642
7	38	100.0	8	19	AA1980296
8	38	100.0	8	19	AA1970375
9	38	100.0	8	19	AA1968308

10	38	100.0	8	19	AA1968365
11	38	100.0	8	19	AA1968700
12	38	100.0	8	19	AA194265
13	38	100.0	8	19	AA192838
14	38	100.0	8	20	AA1942307
15	38	100.0	8	20	AA196871
16	38	100.0	8	20	AA1903780
17	38	100.0	8	20	AA199480
18	38	100.0	8	20	AA1967585
19	38	100.0	8	21	AA193763
20	38	100.0	8	21	AA1929465
21	38	100.0	8	21	AA1936484
22	38	100.0	8	21	AA193785
23	38	100.0	8	21	AA196811
24	38	100.0	8	21	AA199401
25	38	100.0	8	21	AA192564
26	38	100.0	8	22	AA192965
27	38	100.0	8	22	AA197871
28	38	100.0	8	22	AA192562
29	38	100.0	8	22	AA193119
30	38	100.0	8	22	AA192145
31	38	100.0	8	22	AA1909514
32	38	100.0	8	22	AA193855
33	38	100.0	8	22	AA196422
34	38	100.0	8	22	AA1905398
35	38	100.0	8	22	AA1906033
36	38	100.0	8	22	AA194316
37	38	100.0	8	22	AA199354
38	38	100.0	8	22	AA192176
39	38	100.0	8	22	AA191122
40	38	100.0	8	22	AA192065
41	38	100.0	8	22	AA192374
42	38	100.0	8	22	AA1948950
43	38	100.0	8	23	AA1928959
44	38	100.0	8	23	AA1936368
45	38	100.0	8	23	AA1979933
46	38	100.0	8	23	ABG93028
47	38	100.0	8	23	AA1925400
48	38	100.0	8	23	ABG31661
49	38	100.0	8	23	ABG31967
50	38	100.0	8	23	AA199718
51	38	100.0	8	23	AB1908108
52	38	100.0	8	23	AB191273
53	38	100.0	8	23	AA192531
54	38	100.0	8	23	AB196050
55	38	100.0	8	23	AA191945
56	38	100.0	8	23	AB1909907
57	38	100.0	8	23	AA1976942
58	38	100.0	8	23	AA1976802
59	38	100.0	8	23	AA1976869
60	38	100.0	8	23	AA191866
61	38	100.0	8	23	AA191239
62	38	100.0	8	23	AA1909820
63	38	100.0	8	23	AA193215
64	38	100.0	8	23	AA193436
65	38	100.0	8	23	AA1975056
66	38	100.0	8	24	AB1907743
67	38	100.0	8	24	AB1908619
68	38	100.0	8	24	AB1957401
69	38	100.0	8	24	AB1958359
70	38	100.0	8	24	AB1973081
71	38	100.0	8	24	AB1956760
72	38	100.0	8	24	AB1960027
73	38	100.0	8	24	AB1911029
74	38	100.0	9	22	AA1984323
75	38	100.0	9	22	AB1957402
76	38	100.0	10	18	AA1904643
77	38	100.0	10	18	AA1904644
78	38	100.0	10	18	AA1909821
79	38	100.0	10	23	AA1909825
80	38	100.0	12	18	AA1914122
81	38	100.0	12	23	AA1909822
82	38	100.0	12	23	AA1909826

MHC binding peptid
Ovalbumin peptide
Ovalbumin peptide
Ovalbumin artifact
Ovalbumin-derived
OVA peptide sequen
Ovalbumin peptide
Ovalbumin-derived
T-cell activation
Peptide fragment f
Negative control p
Tumour associated
OVA-derived peptid
Altered MHC determ
Ovalbumin protein
Murine ovalbumin M
Altered MHC determ
Ovalbumin 257-264
Cytotoxic T-cell e
Ovalbumin (OVA)-de
Murine ovalbumin (
Human ovalbumin pe
Amino acid sequenc
Chicken ovalbumin
Peptide released f
Chicken ovalbumin
Peptide used to pr
Ovalbumin cytotoxi
Immunodominant CTL
Chicken ovalbumin
Ovalbumin-derived
Miscellaneous pept
Ovalbumin MHC clas
Chicken ovalbumin
Ovalbumin CTL epit
Ovalbumin T-cell e
Mouse class I MHC
Ovalbumin peptide
Chicken ovalbumin
Chicken ovalbumin
Mouse MHC class I
Chicken ovalbumin
Chicken OVA 257-26
Ovalbumin peptide
Cytotoxic T-cell e
Ovalbumin peptide
OVA peptide (257-2
MHC class I-restri
OVA peptide fragme
Ovalbumin driven c
Immunodominant Kb-
Ovalbumin-derived
Hen egg ovalbumin
Chicken ovalbumin
Ovalbumin antigeni
Chicken ovalbumin
Ovalbumin (OVA) re
Synthetic 8mer pep
Ovalbumin-derived
MHC class I peptid
Ovalbumin derived
Ovalbumin immunodo
Ovalbumin immunodo
Peptide used to pr
Synthetic 9mer pep
Ovalbumin-derived
Ovalbumin-derived
Modified ovalbumin
Modified ovalbumin
OVA protein derive
Modified ovalbumin
Modified ovalbumin

83 38 100.0 12 23 AAU09827 Modified ovalbumin
 84 38 100.0 14 23 ABB76049 Peptide insert in
 85 38 100.0 14 23 AAU09823 Modified ovalbumin
 86 38 100.0 14 23 AAU09828 Modified ovalbumin
 87 38 100.0 15 23 AAU09824 Modified ovalbumin
 88 38 100.0 16 24 ABP57403 Synthetic 16mer pe
 89 38 100.0 19 18 AAW19957 Bip-binding domain
 90 38 100.0 19 18 AAW19956 OVA-Bip-binding do
 91 38 100.0 19 23 AAU13446 Chicken MHC class
 92 38 100.0 19 23 AAU13447 Chicken MHC class
 93 38 100.0 19 24 ABP57404 Synthetic 19mer pe
 94 38 100.0 24 14 AAR32294 Sequence of synthe
 95 38 100.0 24 14 AAR32294 Antigenic peptide
 96 38 100.0 24 18 AAW04645 Ovalbumin-derived
 97 38 100.0 24 22 AAG65170 Ovalbumin based pe
 98 38 100.0 24 22 AAB74439 Myelin basic prote
 99 38 100.0 24 23 AEG31664 Ovalbumin (OVA) pe
 100 38 100.0 26 24 ABP57405 Synthetic 26mer pe
 101 38 100.0 26 24 ABP57406 Synthetic 26mer* p
 102 38 100.0 30 23 AAU13448 Chicken MHC class
 103 38 100.0 31 24 ABP57407 Synthetic 31mer pe
 104 38 100.0 35 18 AAW04646 Ovalbumin-derived
 105 38 100.0 36 24 AAO26741 Chicken ovalbumin
 106 38 100.0 43 22 AAB84325 Amino acid sequenc
 107 38 100.0 47 22 AAB84321 Amino acid sequenc
 108 38 100.0 48 22 AAB84322 Tns-DICE ovalbumin
 109 38 100.0 49 22 AAB48953 DICE-I ovalbumin M
 110 38 100.0 57 22 AAB48954 Chicken ovalbumin
 111 38 100.0 100 23 AAE13458 Chicken ovalbumin
 112 38 100.0 100 23 AAE13460 Chicken ovalbumin
 113 38 100.0 103 23 AAE13459 Chicken ovalbumin
 114 38 100.0 103 23 AAE13461 Chicken ovalbumin
 115 38 100.0 106 17 AAB89966 Polytope sequence.
 116 38 100.0 108 23 AAE13462 Chicken ovalbumin
 117 38 100.0 111 23 AAE13463 Chicken ovalbumin
 118 38 100.0 132 21 AAY52575 Amino acid sequenc
 119 38 100.0 386 23 AAE13435 Amino acid sequenc
 120 38 100.0 409 22 AAE31545 Amino acid sequenc
 121 38 100.0 479 22 AAE13112 Human HER300-rGM-C
 122 38 100.0 541 23 AAU99725 Yeast/mouse SS-OVA
 123 38 100.0 564 22 AAE13110 Human HER500 fusio
 124 38 100.0 697 22 AAE13111 Human HER500-rGM-C
 125 38 100.0 948 22 AAB31611 Amino acid sequenc

ALIGNMENTS

RESULT 1
 AAR57996
 ID AAR57996 standard; Protein; 8 AA.
 XX
 AC
 XX
 DT 25-MAR-2003 (updated)
 DT 30-MAR-1995 (first entry)
 XX
 XX
 DE Ova257-264.
 XX
 XX Ova; ovalbumin; cytosol; cytolytic immune response;
 KW vaccinia virus; promoter.
 XX
 OS Synthetic.
 OS
 XX WO9417816-A1.
 XX
 PD 18-AUG-1994.
 XX
 PF 27-JAN-1994; 94WO-US01183.
 XX
 PR 10-FEB-1993; 93US-0016066.
 XX
 XX (DAND) DANA FARBER CANCER INST INC.
 PA

(HARD) HARVARD COLLEGE.
 Goldberg AL, Rock KL;
 WPI; 1994-279383/34.
 Method for blocking cytolytic immune responses - useful for
 treatment of autoimmune diseases and preventing organ and graft
 rejection
 Example 2; Page 44; 89pp; English.
 Ova257-264 was constructed by inserting a synthetic oligonucleotide
 (AAQ67342) behind the vaccinia virus p7.5 early/late promoter in pSc11
 which was modified such that the restriction sites SalI and NotI
 were substituted for the SmaI site. The oligonucleotide consists of
 a SalI site, Kozak's consensus sequence for efficient translation,
 an initiation codon, nucleotides encoding the peptide given in
 AAR57996, two stop codons, and a NotI site.
 (Updated on 25-MAR-2003 to correct FN field.)
 XX Sequence 8 AA;
 SQ

Query Match 100.0%; Score 38; DB 15; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0;
 QY 1 SIINFEKL 8
 |||||
 Db 1 SIINFEKL 8

RESULT 2
 AAR83938
 ID AAR83938 standard; peptide; 8 AA.
 XX
 AC AAR83938;
 XX
 DT 05-JUN-1996 (first entry)
 XX
 DE MHC class I restricted antigenic peptide #8.
 XX
 KW MHC class I; antigen; MAGE; melanoma; breast cancer; bladder cancer;
 KW Titermax; cytotoxic T-lymphocyte; tumour; pathogenic disease; bacteria;
 XX parasite; human; animal.
 OS Synthetic.
 OS
 XX WO9528958-A1.
 XX
 PD 02-NOV-1995.
 XX
 PF 21-APR-1995; 95WO-US04975.
 XX
 PR 22-APR-1994; 94US-0233496.
 XX
 PA (SLOK) SLOAN KETTERING INST CANCER RES.
 XX
 PI Dyall R, Nikolic-Zugic J;
 XX
 DR WPI; 1995-382848/49.
 XX
 XX Cytotoxic T-cell induction by MHC class I-restricted peptide in
 PT adjuvant - useful for treating tumours and bacterial or parasitic
 PT pathogenic diseases
 XX
 PS Claim 11; Page 38; 50pp; English.
 XX
 CC The sequences given in AAR83931-49 are MHC class I restricted 8-12
 CC amino acid antigenic peptides. This peptide represents ovalbumin
 CC residues 257-264. These peptides may be administered to a subject
 CC in combination with a suitable adjuvant, pref. Titermax (RTM), to
 CC induce cytotoxic T-lymphocytes. This method may be used in the

CC treatment of a tumour or a pathogenic disease, esp. diseases of
 CC bacterial or parasitic origin, in humans and animals, e.g. monkeys,
 CC dogs, cows, horses, etc.

XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 16; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 DB 1 SIINFEXL 8

RESULT 3
 AAR89157
 ID AAR89157 standard; peptide; 8 AA.
 XX AC AAR89157;
 XX DT 25-MAR-2003 (updated)
 XX DT 03-SEP-1996 (first entry)
 XX DE Peptide Ova8 used in an MHC stripping/reloading method.
 XX KW Antigen; major histocompatibility complex; cell surface; stimulation;
 KW cytotoxic T lymphocyte; CTL; endogenous; exogenous; peptide; spleen;
 KW allelic restriction; peripheral blood lymphocytes; ganglion; placenta;
 KW native form; infection; tumour; autoimmune disease.
 XX OS Synthetic.
 XX PN WO9601891-A1.
 XX PD 25-JAN-1996.
 XX PP 06-JUL-1995; 95WO-FR00907.
 XX PR 07-JUL-1994; 94FR-0008427.
 XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX PA (INSP) INST PASTEUR.
 XX PI Langlade Demoyen P, Kourilsky P, Abastado J;
 XX WPI; 1996-097621/10.
 XX PT Cell population with high surface density of exogenous peptide bound
 PT to MHC molecules - prepd. by stripping endogenous peptide and
 PT reloading exogenous peptide(s), useful for stimulating cytotoxic
 PT lymphocytes in cases of infection, tumour and auto-immune disease
 XX PS Disclosure; Page 9; 37pp; French.
 XX CC Peptides AAR89154-73 are examples of exogenous peptides which are
 CC "loaded" onto the antigen-presenting major histocompatibility complex
 CC (MHC) on the cell surface in a novel method of stimulating cytotoxic T
 CC lymphocytes. The method involves treating cells, e.g. at a pH 5 or below
 CC or pH 9 or above, to remove the endogenous peptides from the MHC,
 CC followed by recharging the complexes with specified exogenous peptides
 CC having the same allelic restriction as the MHC. The cells are pref.
 CC peripheral blood lymphocytes, spleen, ganglion or placental cells which
 CC are able to present the exogenous antigen in a native form. Recharged
 CC cells contain exogenous peptides at a higher density than native cells
 CC and are used to stimulate specific cytotoxic T lymphocytes in response to
 CC infections, tumours or autoimmune diseases. This peptide is derived from
 CC ovalbumin and belongs to the allelic restriction Kb.
 CC (Updated on 25-MAR-2003 to correct PI field.)
 XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 17; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 DB 1 SIINFEXL 8

RESULT 4
 AAW19955
 ID AAW19955 standard; Peptide; 8 AA.
 XX AC AAW19955;
 XX DT 10-NOV-1997 (first entry)
 XX DB Chicken OVA-peptide.
 XX KW Vaccine; immunotherapy; heat shock protein; OVA; cancer;
 KW infectious disease.
 XX OS Gallus sp.
 XX PN WO9706821-A1.
 XX PD 27-FEB-1997.
 XX PP 16-AUG-1996; 96WO-US13363.
 XX PR 18-AUG-1995; 95US-0002490.
 XX PR 18-AUG-1995; 95US-0002479.
 XX PA (SLOK) SLOAN KETTERING INST CANCER RES.
 XX PI Hartl FU, Hoe MH, Houghton A, Mayhew M, Rothman JE;
 PI Takeuchi Y;
 XX WPI; 1997-165035/15.
 XX PT Compens. for inducing immune response contg. antigen and heat shock
 PT protein - also new hybrid peptide and related nucleic acid, for
 PT treatment of infectious diseases and tumours
 XX PS Example 1; Page 17; 58pp; English.
 XX CC Chicken OVA-peptide (AAW19955) is used in novel hybrid peptides,
 CC OVA-Bip (AAW19956) and Bip-OVA (AAW19957) with heat shock protein (HSP)
 CC Bip binding domain (see also AAW19951). The hybrid protein is
 CC combined in vitro with a HSP, such as hsp70, to form a complex
 CC that, when administered to a subject, induces an immune response.
 CC Vaccine compositions were prepd. by combining recombinant mouse
 CC hsp70, recombinant human hsp40 and Ova-peptide. Combinations of
 CC antigen with hsp70 or a mixture of hsp70 and hsp40 were effective
 CC to produce a cytotoxic T lymphocyte response.
 XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 18; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEXL 8
 |||||
 DB 1 SIINFEXL 8

RESULT 5
 AAW14087
 ID AAW14087 standard; peptide; 8 AA.
 XX AC AAW14087;
 XX DT 20-OCT-1997 (first entry)

XX DE MHC class I molecule binding OVA peptide.
 XX KW Major histocompatibility complex; MHC; target; binding; tumour;
 XX KW cancer; neoplasia; LSTRA; EL-4; identification; detection; screening;
 XX KW tissue typing; Bcr-abl.
 XX OS Mus sp.
 XX PN WO9641188-A1.
 XX PD 19-DEC-1996.
 XX PF 07-JUN-1996; 96WO-US09680.
 XX PR 07-JUN-1995; 95US-0485610.
 XX PA (UNIW) UNIV WASHINGTON.
 XX PI Cheever MA, Chen W;
 XX WPI; 1997-108657/10.
 XX DR Identifying major histocompatibility complex class I binding mols. -
 PT using peptide(s) having a core of 7-14 amino acids with extra amino
 PT acids and a reporter gp. at the N- or C-terminus, useful for tissue
 PT typing
 XX Example 2; Page 19; 41pp; English.
 XX CC AAW14087-91 are peptides derived from LSTRA and EL-4 tumours of Balb/c
 CC mice. The peptides were tested for MHC specificity to find MHC I
 CC specific peptides. These peptides are useful for tissue typing or for
 CC screening for molecules that interact with MHC class I molecules. MHC
 CC class I molecules can be identified using the peptides and also the
 CC peptides are useful in vaccines against disease and infection e.g.
 CC caused by viruses, bacteria or tumours.
 XX Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 18; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 Db |||||
 1 SIINFEKL 8
 RESULT 6
 AAW04642
 ID AAW04642 standard; peptide; 8 AA.
 XX AC AAW04642;
 XX DT 01-AUG-1997 (first entry)
 XX DE Ovalbumin-derived activated CD8+ T cells epitope OVA8.
 XX KW Macrophage; artificial antigen presenting cell; APC; cancer;
 KW tumours; neoplasia; viral infection; retroviral infection;
 KW autoimmune.
 XX OS Synthetic.
 XX PN WO9637107-A1.
 XX PD 28-NOV-1996.
 XX PF 22-MAY-1996; 96WO-US07436.
 XX PR 23-MAY-1995; 95US-0447761.
 XX

PA (SCRI) SCRIPPS RES INST.
 XX DeBruijn MLH, Jackson MR, Peterson PA;
 XX WPI; 1997-020850/02.
 XX DR Prodn. of activated CD8+ T cells directed to specific antigen - can
 XX PT specifically kill target cells useful to treat, e.g. cancer
 XX PS Example 1; Page 26; 84pp; English.
 XX CC The method for the production of activated CD8+ T cells specifically
 CC directed towards a particular antigen involves affixing peptides
 CC corresponding to the particular antigen to an artificial support;
 CC contacting macrophages with the affixed peptides for a time sufficient
 CC for the peptides to be engulfed, and at least a portion of the peptides
 CC to be presented on the surface of the macrophage; and contacting
 CC unprimed CD8+ T cells with the peptide presenting macrophages for a
 CC time sufficient to activate the unprimed CD8+ T cells. The present
 CC sequence represents a peptide designated OVA8 which corresponds to
 CC ovalbumin, a Kb-restricted peptide antigen. This is the optimal
 CC peptide. Small extensions to the optimal peptide affect the affinity
 CC of the peptide for soluble class I molecules in vitro e.g. the addition
 CC of two amino acids to the amino-terminus lowers the affinity to Kb by
 CC 76-fold compared to the optimal peptide; addition of two amino acids to
 CC the carboxy-terminus lowers the affinity by 4-fold. The method,
 CC macrophages and artificial antigen presenting cell, having a peptide
 CC corresponding to the particular antigen present on its surface and at
 CC least a portion of an artificial support in its interior, can be used to
 CC treat conditions (e.g. cancer, tumours, neoplasia, viral or retroviral
 CC infection or autoimmune or autoimmune-type conditions) in patients via
 CC the specific killing of target cells.
 XX Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 18; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 Db |||||
 1 SIINFEKL 8
 RESULT 7
 AAW80296
 ID AAW80296 standard; peptide; 8 AA.
 XX AC AAW80296;
 XX DT 08-JAN-1999 (first entry)
 XX DE Amino acids 257-364 of chicken ovalbumin used as an antigen.
 XX KW Antisense oligonucleotide; antigen processing protein; TAP;
 KW transporter; proteasome; antigen-presenting cell; cancer; infection;
 KW cytotoxic T cell; chicken ovalbumin.
 XX OS Synthetic.
 XX OS Gallus sp.
 XX PN US5831068-A.
 XX PD 03-NOV-1998.
 XX PF 20-AUG-1996; 96US-0700035.
 XX PR 20-AUG-1996; 96US-0700035.
 XX PR 21-AUG-1995; 95US-0517373.
 XX PA (UYDU-) UNIV DUKE.
 XX PI Gilboa E, Nair SK;

Best Local Similarity 100.0%; Pred. No. 9.3e+05; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
|||||||
DB 1 SIINFEKL 8

RESULT 10
AAW68365
ID ID AAW68365 standard; peptide; 8 AA.
XX AC AAW68365;
XX AC
XX 25-MAR-2003 (updated)
XX 14-OCT-1998 (first entry)
XX MHC binding peptide from ovalbumin.
XX DE
XX Antigen; major histocompatibility complex; MHC; lymphocyte; detection;
KW immobilisation; cytotoxic T-cell; tumour; leukaemia; lymphoma;
KW viral infection.
XX
XX Synthetic.
OS
XX WO9744667-A2.
PN
XX 27-NOV-1997.
PD
XX
XX 21-MAY-1997; 97WO-FR00892.
PP
XX 21-MAY-1996; 96US-0651925.
XX
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PA (INSP) INST PASTEUR.
PA
XX
XX Langladedemoyen P, Lone Y, Kourilsky P, Abastado J;
PI
XX
XX WPI; 1998-018653/02.
XX
XX
XX Detection, purification and elimination of antigen-specific
PT lymphocytes - for producing cytotoxic T cells for immuno-therapy of
PT cancers and viral infection
PT
XX
XX Disclosure; Page 29; 222pp; French.
PS
XX
XX Peptides AAW68301-W68384 are examples of antigens (Ag) which can be
CC loaded onto recombinantly produced major histocompatibility complex
CC (MHC) molecules in a method of detecting antigen-specific lymphocytes.
CC The MHC-antigen complex is then immobilised on a solid support and a
CC sample containing cells recognising the MHC-Ag complex may be isolated.
CC This peptide is derived from amino acids 258-276 of ovalbumin. A
CC similar method is used to isolate, purify or eliminate Ag-specific
CC T-cells or to produce Ag-specific cytotoxic T-cells (CTC). The method is
CC also used to detect and quantify tumour-specific T-cells and to generate
CC CTC for specific killing of tumour cells (solid tumours, leukaemia or
CC lymphoma) by injection into a human or animal, but also for treating
CC viral infections.
CC
XX (Updated on 25-MAR-2003 to correct PI field.)
XX

Sequence 8 AA;
QY Query Match 100.0%; Score 38; DB 19; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
|||||||
DB 1 SIINFEKL 8

RESULT 11
AAW60700

```

XX DE Ovalbumin peptide OVA.
XX KW Pep-MHC complex; cytotoxic; T cell; cancer; ovarian; brain;
XX KW major histocompatibility complex.
XX OS Synthetic.
XX PN WO9807441-A1.
XX PD 26-FEB-1998.
XX PF 22-AUG-1997; 97WO-US14814.
XX PR 23-AUG-1996; 96US-0023437.
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX PA (UNII ) UNIV ILLINOIS FOUND.
XX PI Eisen HN, Kranz DM;
XX DR WPI; 1998-168897/15.
XX CC Composition for targeting an allo-reactive response to specific
XX PT cells - comprises MHC-peptide complex bound to specific targeting
XX PT agent, the complex having at least one allogeneic component,
XX PT specifically for treating ovarian and brain cancers
XX PS Examples; Page 26; 45pp; English.
XX CC Peptide OVA derived from ovalbumin was used in the production of soluble
XX CC pep-MHC complexes. The pep-MHC (peptide-major histocompatibility
XX CC complex) complexes, including both MHC heavy and light chains, bound to a
XX CC specific targeting molecule can be used to target a cytotoxic T cell
XX CC response to specific cells. This is particularly useful for targeting
XX CC cancer cells, specifically of the ovary or brain. The method can also
XX CC be used to eliminate an entire cell type for example during bone marrow
XX CC therapy.
XX SQ Sequence 8 AA;
XX Query Match 100.0%; Score 38; DB 19; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 SIINFEKL 8
XX DB |||||
XX 1 SIINFEKL 8
XX RESULT 13
XX AAWS2838
XX ID AAWS2838 standard; peptide; 8 AA.
XX AC AAWS2838;
XX XX 24-JUN-1998 (first entry)
XX DE Ovalbumin artificial target antigen.
XX KW Ovalbumin; antigen; ATA; cytotoxic T lymphocyte; CTL; tumour; prevention.
XX OS Gallus sp.
XX PN WO9800163-A1.
XX PD 08-JAN-1998.
XX PF 18-JUN-1997; 97WO-US10195.
XX PR 28-JUN-1996; 96US-0675332.
XX PA (DAND ) DANA FARBER CANCER INST INC.

```

```

PA (UYEI-) UNIV PITTSBURGH.
XX PI Palo LD, Rock KL;
XX DR WPI; 1998-086733/08.
XX PT Inducing anti-tumour cytotoxic T lymphocytes - by cross-priming
XX PT with artificial antigen, then immunisation with tumour cells
XX PT modified in vitro to express the same antigen, does not require
XX PT characterisation of tumour-specific antigens
XX PS Example; Page 21; 44pp; English.
XX CC Ovalbumin can be used as an artificial target antigen (ATA) to
XX CC promote a cytotoxic T lymphocyte mediated response in mammals. Tumour
XX CC cells from the host can be engineered to include ATA and therefore
XX CC induce anti-tumour cytotoxic T Lymphocytes. This method can be used for
XX CC the treatment and prevention of a wide range of tumours even when
XX CC the tumour is inaccessible or where metastases are being targeted.
XX SQ Sequence 8 AA;
XX Query Match 100.0%; Score 38; DB 19; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 SIINFEKL 8
XX DB |||||
XX 1 SIINFEKL 8
XX RESULT 14
XX AA42307
XX ID AA42307 standard; peptide; 8 AA.
XX AC AA42307;
XX XX 06-DEC-1999 (first entry)
XX DE Ovalbumin-derived peptide antigen.
XX KW Immunity; human leukocyte antigen; HLA; MHC; antigen;
XX KW major histocompatibility complex; presentation; solubility;
XX KW dendritic cell.
XX OS Synthetic.
XX PN WO9947646-A1.
XX PD 23-SEP-1999.
XX PF 19-MAR-1999; 99WO-US06627.
XX PR 20-MAR-1998; 98US-0078832.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PA (BIOP-) BIOPOL SCRL.
XX PI Rescigno M, Girolomoni G, Corinti S, Ricciardi-Castagnoli P;
XX DR WPI; 1999-571834/48.
XX PT Preparation of dendritic cells which present antigens, used for
XX PT stimulating an immune response by immunocompetent cells
XX PS Example 6; Page 13; 43pp; English.
XX CC This sequence represents an ovalbumin-derived peptide antigen which
XX CC is presented on major histocompatibility complex (MHC) Class I molecules
XX CC of dendritic cells via the use of a novel process to improve soluble
XX CC protein antigen presentation. This process uses dendritic cells which
XX CC have internalised bacterial cells recombinantly expressing ovalbumin. The
XX CC dendritic cells which have internalised the bacteria can be used to

```

CC generate an immune response which is stronger than the response which
CC would be generated when soluble antigen alone is used. The dendritic
CC cells which result from the internalisation are characterised by MHC
CC Class I and Class II molecules which have much longer half lives and
CC greater stability than comparable dendritic cells which have not
CC internalised such bacteria. The dendritic cells can be contacted with
CC immunocompetent cells for stimulating an immune response. The dendritic
CC cells can also be used to stimulate maturation of an immature dendritic
CC cell.

XX SQ Sequence 8 AA;
Query Match 100.0%; Score 38; DB 20; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
|||
Db 1 SIINFEKL 8

RESULT 15
AAV16871
ID AAV16871 standard; peptide; 8 AA.
XX AAV16871;
AC AAV16871;
XX 20-JUL-1999 (first entry)
DT 20-JUL-1999 (first entry)
XX OVA peptide sequence.
DE OVA peptide sequence.
XX Conjugate peptide; heat shock protein; hsp; phage display library; virus;
KW surface protein; tethering peptide; chaperone process; cytokine; cancer;
KW neoplastic disease; infectious disease; bacterium; immune system; fungus;
KW acquired immune deficiency; autoimmune disease.
XX Synthetic.
OS Synthetic.
XX WO9922761-A1.
PN WO9922761-A1.
XX 14-MAY-1999.
PD 14-MAY-1999.
XX 22-OCT-1998; 98WO-US22335.
PF 22-OCT-1998; 98WO-US22335.
XX 31-OCT-1997; 97US-0961707.
PR 31-OCT-1997; 97US-0961707.
XX (SLOAN) SLOAN KETTERING INST CANCER RES.
PA (SLOAN) SLOAN KETTERING INST CANCER RES.
XX Hartl U, Hoe MH, Houghton A, Mayhew M, Moroi Y;
PI Ouerfelli O, Rothman JE;
PI Ouerfelli O, Rothman JE;
XX WPI; 1999-313177/26.
DR WPI; 1999-313177/26.
XX Identifying peptides which bind heat shock proteins
PT Identifying peptides which bind heat shock proteins
XX Examples; Page 49; 155pp; English.
PS Examples; Page 49; 155pp; English.
XX The invention relates to conjugate peptides engineered to noncovalently
CC bind to heat shock proteins (hsp). A method of identifying a hsp binding
CC peptide comprises (a) contacting a phage display library having
CC bacteriophage expressing, in a surface protein, inserted peptides with a
CC hsp target, and bound to a benzoguinone ansamycin antibiotic (BAA), in a
CC physiologic binding buffer; (b) isolating a phage binding to the hsp
CC target; and (c) identifying the inserted peptide expressed. The peptides
CC which bind to a hsp can be used as tethering peptides for a hsp which may
CC serve as an accessory in a chaperone process and/or may comprise a
CC cytokine. They can also be coupled to antigens to induce an immune
CC response. Such compositions can be used for treating neoplastic disease,
CC e.g. cancers, infectious diseases, e.g. diseases caused by a bacterium,
CC virus, protozoan, mycoplasma, fungus, yeast, parasite or prion, or a
CC disease of the immune system, e.g. acquired immune deficiencies or
CC autoimmune diseases.

XX SQ Sequence 8 AA;
Query Match 100.0%; Score 38; DB 20; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
|||
Db 1 SIINFEKL 8

RESULT 16
AAV03780
ID AAV03780 standard; peptide; 8 AA.
XX AAV03780;
AC AAV03780;
XX 23-JUN-1999 (first entry)
DT 23-JUN-1999 (first entry)
XX Ovalbumin peptide fragment (residues 257-264).
DE Ovalbumin peptide fragment (residues 257-264).
XX Dendritic cell-derived factor; proliferation; interferon gamma; IFNgamma;
KW T cell; granulocyte macrophage colony-stimulating factor; GM-CSF;
KW T cell stimulatory factor; lymphocyte; interleukin-2; lipopolysaccharide;
KW autoimmune response; inflammatory; ovalbumin.
XX Synthetic.
OS Synthetic.
XX Gallus gallus.
XX WO9918909-A2.
PN WO9918909-A2.
XX 22-APR-1999.
PD 22-APR-1999.
XX 14-OCT-1998; 98WO-US21614.
PF 14-OCT-1998; 98WO-US21614.
XX 14-OCT-1997; 97US-0062405.
PR 14-OCT-1997; 97US-0062405.
XX (LUDW-) LUDWIG INST CANCER RES.
PA (LUDW-) LUDWIG INST CANCER RES.
XX Dunn A, Marino MW, Noguchi Y, Old LJ, Wada H;
PI Dunn A, Marino MW, Noguchi Y, Old LJ, Wada H;
PI Dunn A, Marino MW, Noguchi Y, Old LJ, Wada H;
XX WPI; 1999-277418/23.
DR WPI; 1999-277418/23.
XX Dendritic cell and T cell derived factors for regulation of T cell
PT Dendritic cell and T cell derived factors for regulation of T cell
XX proliferation and interferon gamma production
PS proliferation and interferon gamma production
XX Examples; Page 37; 68pp; English.
PS Examples; Page 37; 68pp; English.
XX The invention relates to a dendritic cell-derived factor that restores
CC proliferation and interferon gamma (IFNgamma) production to T cells from
CC granulocyte macrophage colony-stimulating factor (GM-CSF). The dendritic
CC cell-derived factor which is a T cell stimulatory factor modulates the
CC effect of GM-CSF on production of IFNgamma by lymphocytes and the
CC response of GM-CSF -/- T cells to interleukin-2, and corrects the
CC lipopolysaccharide-induced defect in IFNgamma production. The dendritic
CC cell-derived factor is used in vivo to increase proliferation of T cells
CC and/or IFNgamma production, e.g. during immunisation to increase the
CC response to an antigen. Agents with T cell derived factor activity are
CC used to decrease production of IFNgamma, e.g. to reduce autoimmune
CC responses and inflammatory reactions. Agents that bind to the factors,
CC e.g. antisense sequences, are used to treat excessive/inadequate T cell
CC proliferation and IFNgamma production. The dendritic cell-derived factor
CC and T cell factor (or their fragments) can also be used to raise
CC antibodies and as components in immunoassay and diagnostic systems, also
CC for in vitro studies. The present sequence represents an ovalbumin
CC peptide fragment used in the course of the invention.

XX SQ Sequence 8 AA;
Query Match 100.0%; Score 38; DB 20; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
|||
Db 1 SIINFEKL 8


```

QY      1 SIINFEKL 8
DB      1 SIINFEKL 8

RESULT 17
AAW99480
ID AAW99480 standard; peptide; 8 AA.
XX
AC AAW99480;
XX
DT 08-JUN-1999 (first entry)
XX
DE Ovalbumin-derived peptide OVA.
XX
KW Matrix protein; conjugate; mutant; major histocompatibility complex; MHC;
KW class I molecule; beta-2-microglobulin; stimulation; immunity; tumour.
XX
OS Synthetic.
XX
PN WO9911775-A1.
XX
PD 11-MAR-1999.
XX
PF 20-AUG-1998; 98WO-US17308.
XX
PR 29-AUG-1997; 97US-0920413.
XX
PA (GEO) GEN HOSPITAL CORP.
PA (HARD) HARVARD COLLEGE.
XX
PI Garboczi DN, Walker J;
XX
DR WPI; 1999-205182/17.
XX
PT Method for conjugating a mutant major histocompatibility complex
PT class I molecule and a compound - useful for stimulating immunity in
PT an individual, and eradicating undesired cells, especially tumours
XX
PS Example III-4; Page 22; 37pp; English.
XX
CC The invention relates to the preparation of a conjugate of a mutant major
CC histocompatibility complex (MHC) class I molecule (containing a Tyr67Cys
CC amino acid substitution in the beta2-microglobulin subunit) and a
CC compound. This sequence corresponds to the ovalbumin-derived peptide
CC OVA. The peptide is used as a control to prepare a hybrid peptide-MHC
CC class I tetramer in which a mutant human beta2-microglobulin subunit
CC binds a mouse MHC class I molecule. The conjugates are useful for
CC stimulating immunity in an individual, and eradicating undesired cells
CC (e.g. tumours).
XX
SQ Sequence 8 AA;
Query Match 100.0%; Score 38; DB 20; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIINFEKL 8
DB      1 SIINFEKL 8

RESULT 18
AAW67585
ID AAW67585 standard; peptide; 8 AA.
XX
AC AAW67585;
XX
DT 02-MAR-1999 (first entry)
XX
DE T-cell activation peptide #3.
XX
KW Activated T helper cell; CD4+; cytotoxic T cell; CD8+; liposome; epitope;

peripheral blood lymphocyte; antigen-presenting cell; APC; virus; tumour;
bacterium; parasite; cytokine; vaccine; cancer; malaria; HIV; hepatitis;
tuberculosis.
OS Synthetic.
XX
PN WO980527-A1.
XX
PD 12-NOV-1998.
XX
PF 07-MAY-1998; 98WO-US09288.
XX
PR 08-MAY-1997; 97US-0045949.
XX
PA (BIOM-) BIOMIRA INC.
XX
PI Agrawal B, Krantz MJ, Longenecker BM, Reddish MA;
XX
DR WPI; 1999-034715/03.
XX
PT Method of activation of T cells - by exposure to antigen-presenting
PT cells loaded with antigen in liposome, used for, e.g. treating
PT cancer and microbial infections
XX
PS Disclosure; Page 6; 75pp; English.
XX
CC Peptides AAW67583-WG7611 are used to produce activated T helper (CD4+)
CC and cytotoxic (CD8+) T cells. The activated T cells are produced by
CC treating peripheral blood lymphocytes with liposome-encapsulated peptide
CC antigen to generate Ag-loaded antigen-presenting cells (APC), contacting
CC naive or anergic T-cells with these APC, and isolating the resulting
CC activated T-cells. The cells are specific for a particular antigen,
CC particularly one derived from a tumour, but also those from viruses,
CC bacteria and other parasites. It can also be used to identify antigens
CC and epitopes able to generate an Ag-specific T-cell response (by
CC assessing proliferation and cytokine release). Also the Ag-loaded APC
CC can be used as cellular vaccines for treating cancer (claimed) or other
CC diseases (e.g. malaria, human immune deficiency virus infection,
CC hepatitis, tuberculosis). The activated T-cells can be used to treat the
CC same conditions by adoptive T-cell transfer therapy.
XX
SQ Sequence 8 AA;
Query Match 100.0%; Score 38; DB 20; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIINFEKL 8
DB      1 SIINFEKL 8

RESULT 19
AAI3763
ID AAI3763 standard; peptide; 8 AA.
XX
AC AAI3763;
XX
DT 02-FEB-2001 (first entry)
XX
DE Peptide fragment from ovalbumin OVA.
XX
KW T-cell; immune response; antigen; epitope; B7 family molecule;
KW Leukocyte function-associated antigen-3; LFA-3;
KW Intercellular adhesion molecule-1; ICAM-1; vaccine; immunotherapy;
KW colon polyp; Crohn's disease; ulcerative colitis; breast lesion;
KW tumour; ovalbumin.
XX
OS Unidentified.
XX
PN WO200034494-A1.
XX
PD 15-JUN-2000.

```

XX PF 12-NOV-1999; 99WO-US26866.
 XX PR 09-DEC-1998; 98US-0111582.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX PA (THER-) THERION BIOLOGICS CORP.
 XX PI Schlom J, Hodge J, Panicali D;
 XX DR WPI; 2000-431307/37.
 XX CC Novel recombinant vector useful as immunogens and vaccines for
 PT stimulating and enhancing immunological responses to target cells and
 PT antigens expresses multiple co-stimulatory molecules such as B7-1,
 PT LFA-3, ICAM-1 -
 XX CC Example 31; Page 80; 188pp; English.
 XX CC Costimulatory molecules have important roles in T-cell activation and
 CC therefore the immune response. The present invention relates to
 CC recombinant vectors which comprise of foreign nucleic acid sequences
 CC encoding at least three costimulatory molecules: a B7 family molecule,
 CC Leukocyte function-associated antigen-3 (LFA-3, human CD58) and
 CC Intercellular adhesion molecule-1 (ICAM-1, CD54) and optionally a foreign
 CC gene encoding a target antigen or immunological epitope. The present
 CC sequence is one such target antigen used in the present invention. The
 CC present sequence is a tumour-associated antigen. The vector of the
 CC present invention would be useful for providing an enhanced immune
 CC response to the present target antigen. The vector of the present
 CC invention may therefore be useful in immunotherapy for treating or
 CC preventing diseases caused by viruses, bacteria, protozoans, parasites,
 CC premalignant cells and tumour cells. The recombinant vector can be used
 CC to treat or prevent preneoplastic or hyperplastic states such as colon
 CC polyps, Crohn's disease, ulcerative colitis and breast lesions.
 XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 21; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFKEKL 8
 |||||
 Db 1 SIINFKEKL 8
 RESULT 20
 AAB29465
 ID AAB29465 standard; peptide; 8 AA.
 XX AC AAB29465;
 XX DT 09-FEB-2001 (first entry)
 XX DE Negative control peptide used in a cytotoxic T cell assay.
 XX KW Telomerase antigen variant; HLA-A2-binding; class I MHC;
 XX KW human leukocyte antigen; major histocompatibility complex;
 XX KW cytotoxic T-cell response; antigen-presenting cell; APC;
 XX KW telomerase-expressing cell; cancer; anticancer vaccine.
 XX OS Synthetic.
 XX PI WO200061766-A2.
 XX PD 19-OCT-2000.
 XX PF 07-APR-2000; 2000WO-IB00610.
 XX PR 09-APR-1999; 99US-0128539.
 XX PS (BIOM-) BIOMIRA INC.
 XX CC

XX PI Agrawal B, Longenecker BM;
 XX DR WPI; 2000-679493/66.
 XX CC New telomerase-specific T-cell antigens useful for generating T-cell
 PT responses against telomerases and for producing vaccines for treating
 PT or preventing cancer by in vivo or ex vivo techniques -
 XX CC Example 1; Page 23; 34pp; English.
 XX CC The invention relates to a human telomerase peptide antigen (AAB29461)
 CC which binds to a class I HLA (human leukocyte antigen, MHC, major
 CC histocompatibility complex), and to conservatively substituted variants
 CC thereof. The invention also relates to a vaccine comprising a telomerase
 CC antigen or antigen variant, a nucleotide encoding a telomerase antigen
 CC or variant, and a method of producing telomerase-primed antigen-
 CC presenting cell (APC) comprising contacting an APC with a composition
 CC containing a telomerase antigen or variant. The telomerase antigens or
 CC vaccine compositions are useful for inducing a cytotoxic T-cell immune
 CC response against telomerase and hence against telomerase-expressing
 CC cells (i.e., cancer cells. Additionally, the telomerase antigen-primed
 CC APC may be coadministered with interleukin-2 for cancer treatment or
 CC prevention. The present sequence represents a peptide used in the
 CC exemplification of the invention in an assay of the cytotoxic activity of
 CC T-cells.
 XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 21; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFKEKL 8
 |||||
 Db 1 SIINFKEKL 8
 RESULT 21
 AAB26484
 ID AAB26484 standard; peptide; 8 AA.
 XX AC AAB26484;
 XX DT 16-JAN-2001 (first entry)
 XX DE Tumour associated OVA peptide.
 XX KW Immune response; vaccine; cancer; infection; tumour; OVA.
 XX OS Unidentified.
 XX PN WO200050080-A1.
 XX PD 31-AUG-2000.
 XX PF 23-FEB-2000; 2000WO-US04565.
 XX PR 26-FEB-1999; 99US-0261473.
 XX PA (UYDU-) UNIV DUKE.
 XX PI Gilboa E, Nair SK, Nicchitta CV;
 XX DR WPI; 2000-558368/51.
 XX CC Eliciting immune response in vertebrate for prevention and treatment of
 PT cancer and infectious diseases involves administering purified complex
 PT comprising calreticulin bound to an antigenic molecule -
 XX PS Disclosure; Page 7; 82pp; English.
 XX CC The present invention relates to a method of eliciting an immune
 CC

CC response by administering a composition of a purified complex of
 CC calreticulin bound to an antigenic molecule. The present invention is
 CC useful for prevention and treatment of cancer and infectious disease
 CC in a vertebrate especially of humans. The present sequence is the
 CC tumour antigenic peptide OVA which was used in the invention.
 XX
 SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 21; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFELK 8
 |||||
 Db 1 SIINFELK 8
 |||||

RESULT 22

AAB13785
 ID AAB13785 standard; peptide; 8 AA.

XX
 AC AAB13785;

XX
 DT 10-NOV-2000 (first entry)

XX
 DE OVA-derived peptide.

XX Chicken; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
 KW major histocompatibility complex class 1; MHC class 1; antigen; tumour;
 KW prostate; breast; multiple myeloma; OVA peptide.
 XX

OS Gallus domesticus.

XX WO2000035949-A1.

XX
 PD 22-JUN-2000.

XX
 PF 14-DEC-1999; 99WO-US29724.

XX
 PR 14-DEC-1998; 98US-0112324.

XX (DEND-) DENDREON CORP.

XX Laus R, Hakim I, Vidovic D;

XX
 DR WPI; 2000-442365/38.

XX Antigens modified by the covalent addition of a peptide that
 PT facilitates entry into antigen presenting cells, useful for producing
 PT compositions for immunizing against tumors and pathogens -

XX Example 1; Page 10; 34pp; English.

XX The present invention relates to compositions of modified soluble protein
 CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
 CC response i.e. a major histocompatibility complex (MHC) class I molecule
 CC peptide sequence which facilitates entry of the antigen into antigen
 CC presenting cells (APCs). The present sequence is a peptide derived from
 CC the chicken antigen OVA. This peptide was used to prepare the modified
 CC antigens. The modified antigen composition may be used for immunising
 CC against, or treating a tumour e.g. prostate and breast carcinoma or
 CC multiple myeloma, or pathogen in mammals.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 21; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFELK 8
 |||||
 Db 1 SIINFELK 8
 |||||

RESULT 23

AAV68311
 ID AAV68311 standard; Peptide; 8 AA.

XX
 AC AAV68311;

XX
 DT 13-APR-2000 (first entry)

XX
 DE Altered MHC determinant binding peptide SEQ ID NO:143.

XX MHC class I; major histocompatibility complex; microglobulin; antigen;
 KW immune response; immunisation; AIDS; multiple sclerosis; toxic shock;
 KW cancer; lupus erythematosus; snake bite; cytostatic; antiviral;
 KW immunomodulatory; dermatological; immunosuppressive; antiinflammatory;
 KW neuroprotective.

XX Unidentified.

XX US6011146-A.

XX
 PD 04-JAN-2000.

XX
 PF 07-JUN-1995; 95US-0481985.

XX
 PR 15-NOV-1991; 91US-0792473.

XX
 PR 05-DEC-1991; 91US-0801818.

XX (INSP) INST PASTEUR.

XX (INRM) INST NAT SANTE & RECH MEDICALE.

XX Kourilsky P, Mottez E, Abastado J;

XX
 PI WPI; 2000-125951/11.

XX New recombinant DNA encoding covalently linked form of major
 PT histocompatibility complex Class I determinant, used for immune system
 PT stimulation, e.g. for treating cancer -

XX Disclosure; Column 12; 88pp; English.

XX The present invention describes a recombinant DNA molecule (I)
 CC containing a sequence (Ia) that encodes an altered MHC (major
 CC histocompatibility complex) Class I determinant (II) comprises a
 CC polypeptide with alpha1, alpha2, alpha3 and beta2-microglobulin
 CC domains, in which alpha3 and beta2 are covalently linked, thorough C-
 CC and N-termini respectively, via a nucleotide spacer sequence encoding a
 CC polypeptide. (II) includes an antigen-binding site and when (II) and
 CC the antigen are associated they are recognized by a mammalian T cell
 CC receptor (TCR). (I) are used to produce (II) which are used to study
 CC functional interactions between the various MHC domains. They can also
 CC be used to modulate (in vivo or in vitro) the immune system by inducing
 CC an effector response (cytotoxicity, antibody synthesis, phagocytosis)
 CC of immune system cells, typically for treating, or immunising against;
 CC cancer, acquired immune deficiency syndrome, lupus erythematosus,
 CC multiple sclerosis, toxic shock and snake bite, but also for selective
 CC destruction of autoreactive cells, diagnostically to assay T cell
 CC receptors and to raise specific antibodies (useful for diagnosis,
 CC therapy, studying MHC-associated cellular processes and for affinity
 CC purification). AAZ57558 and AAV68186 to AAV68316 are sequences used in
 CC the exemplification of the present invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 21; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFELK 8
 |||||
 Db 1 SIINFELK 8
 |||||

```

RESULT 24
AA52564
ID AAY59401 standard; Protein; 8 AA.
XX AC
XX AAY59401;
XX
XX 17-MAR-2000 (first entry)
XX DT
XX Ovalbumin protein fragment.
XX DE
XX
XX Ovalbumin; chicken; nonvirulent bacterium; cytolysin;
XX KW diagnosis; gene therapy; polypeptide delivery; gene delivery.
XX KW
XX Gallus sp.
XX OS
XX US6004815-A.
XX PN
XX 21-DEC-1999.
XX PD
XX 13-AUG-1998; 98US-0133914.
XX PF
XX 13-AUG-1998; 98US-0133914.
XX PR
XX (REGC ) UNIV CALIFORNIA.
XX PA
XX Portnoy DA, Higgins DB;
XX PI
XX
XX WPI; 2000-072064/06.
XX DR
XX Nonvirulent bacterium useful for the intracellular delivery of agents
XX PT to eukaryotic cells -
XX PT
XX Example; Column 12; 14pp; English.
XX PS
XX This sequence represents a fragment of the chicken ovalbumin.
XX CC The invention relates to a nonvirulent bacterium (1), comprising
XX CC a first gene encoding a nonsecreted foreign functional cytolysin operably
XX CC linked to a heterologous promoter which expresses cytolysin in the
XX CC bacterium and a second gene encoding a different foreign agent. The
XX CC bacterium can be used to deliver a wide variety of foreign agents to
XX CC eukaryotic cells for applications such as diagnosis, therapy including
XX CC prophylactic treatments such as immunisations and gene therapy especially
XX CC for single gene disorders suitable for localised treatment and
XX CC biosynthesis. They can be used to deliver antigenic polypeptides
XX CC presented in association with major histocompatibility (MHC) proteins to
XX CC antigen-presenting cells. The bacterium can be engineered to deliver
XX CC libraries of agents for screening. The foreign agent can be delivered to
XX CC any target cell capable of carrying out endocytosis of the subject
XX CC microbe, in particular epithelial cells, endothelial cells, muscle cells,
XX CC liver cells, pancreatic cells, neural cells, fibroblasts, tumour cells,
XX CC and leukocytes. The foreign agent introduced into the cell is a nucleic
XX CC acid and/or protein which is bioactive in and therapeutic to the target
XX CC eukaryote. Bacterium mediated delivery of protein does not require
XX CC protein purification and high levels of protein can be delivered to the
XX CC cytosol of virtually all cells in culture. The process is rapid and
XX CC efficient and large enzymatically active proteins can be delivered.
XX CC
XX SQ Sequence 8 AA;
XX
Query Match 100.0%; Score 38; DB 21; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEXL 8
| | | | |
DB 1 SIINFEXL 8
| | | | |
RESULT 25
AA52564
ID AAY52564 standard; peptide; 8 AA.
XX

```

```

AC
XX
XX AAY52564;
XX DT 28-FEB-2000 (first entry)
XX DE Murine ovalbumin MHC class I epitope.
XX DE
XX
XX Chimeric; ovalbumin; pan DR epitope; expression vector;
XX KW promoter; major histocompatibility complex; MHC; targeting; peptide;
XX KW epitope; antigen; presentation; class I; cytosolic pathway;
XX KW endoplasmic reticulum; class II; extracellular antigen;
XX KW endocytic pathway; helper T lymphocyte; HTL; universal epitope;
XX KW cytotoxic T lymphocyte; CTL; immune response; immunogenicity; assay;
XX KW vaccine; immunity; infection; pathogen; virus; HIV; HBV; HCV;
XX KW hepatitis B; hepatitis C; Bacterium; protozoan; tumour cell;
XX KW autoimmune disease; activation; antiviral; antimalarial;
XX KW immunoprotective.
XX OS
XX Synthetic.
XX OS Mus sp.
XX
XX PN W0958658-A2.
XX PD 18-NOV-1999.
XX
XX PF 13-MAY-1999; 99WO-US10646.
XX
XX PR 13-MAY-1998; 98US-0078904.
XX PR 15-MAY-1998; 98US-0085751.
XX
XX (EPIN-) EPIMUNE INC.
XX PA
XX Fikes JD, Hermanson GG, Sette A, Ishioka GY, Livingston B;
XX PI Chesnut RW;
XX
XX WPI; 2000-039103/03.
XX DR N-PSDB; AA238683.
XX
XX Expression vectors encoding major histocompatibility targeting
XX PT sequence, used as, e.g. tumor vaccines -
XX
XX Example 1; Page 45; 130pp; English.
XX
XX This sequence represents a murine ovalbumin MHC class I epitope
XX CC encoded by the AOS minigene insert of the expression vector pMIN.0
XX CC (AA238634). This insert encodes several MHC class I epitopes,
XX CC including this sequence, plus DNA encoding the universal MHC
XX CC class II (helper T) epitope, pan DR epitope (PADRE), and was used
XX CC in an exemplification of the present invention. The invention
XX CC relates to a novel expression vector comprising a promoter operably
XX CC linked to a fusion gene encoding a major histocompatibility complex
XX CC (MHC) targeting sequence, and two or more heterologous peptide
XX CC epitopes. The MHC targeting sequence may be a class I targeting
XX CC sequence, which directs an MHC class I epitope to a cytosolic pathway or
XX CC to the endoplasmic reticulum, or an MHC class II targeting sequence,
XX CC which directs extracellular antigens to enter the endocytic pathway to
XX CC be processed into antigen peptides for presentation on MHC class II
XX CC molecules. The heterologous epitopes for presentation on MHC class I
XX CC lymphocyte (HTL) epitopes, or a cytotoxic T lymphocyte (CTL) epitope and
XX CC a universal HTL epitope such as a pan DR epitope (PADRE). The vectors
XX CC are useful for stimulating an immune response in vivo, as well as for use
XX CC in assaying the human immunogenicity of a human T cell peptide epitope
XX CC in vivo in a non-human mammal. They provide a nucleic acid vaccine for
XX CC enhancing immunity against infectious pathogens, such as viruses (e.g.,
XX CC HIV, hepatitis B (HBV) and hepatitis C (HCV)) bacteria, protozoa (e.g.,
XX CC Plasmodium falciparum, the cause of malaria) and also tumour cells and
XX CC autoimmune diseases. Universal MHC class II epitopes are advantageously
XX CC combined with other MHC class I and class II epitopes to increase the
XX CC number of cells that are activated in response to a given antigen and
XX CC provide a broader population coverage of MHC-reactive alleles.
XX
XX SQ Sequence 8 AA;
XX
Query Match 100.0%; Score 38; DB 21; Length 8;

```

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFEKL 8
| | | | |
Db 1 SIINFEKL 8

RESULT 26

AAV52965
ID AAV52965 standard; Peptide; 8 AA.

XX AC AAY52965;

XX DT 14-FEB-2000 (first entry)

XX DE Altered MHC determinant binding peptide SEQ ID NO:143.

XX KW Major histocompatibility complex; MHC class I; MHC class II; antigen;
XX KW immune response; diagnosis; antibody; immunisation; autoimmune disease;
XX KW acquired immune deficiency syndrome; AIDS; cytostatic; dermatological;
XX KW anti-inflammatory; neuroprotective; immunosuppressive; antichyroid;
XX KW vaccine; lupus erythematosus; multiple sclerosis; thyroiditis;
XX KW toxic shock; tumour; snakebite.

XX OS Synthetic.

XX OS Synthetic.

XX PN US5976551-A.

XX PD 02-NOV-1999.

XX PF 07-JUN-1995; 95US-0484905.

XX PR 05-DEC-1991; 91US-0801818.

XX PR 15-NOV-1991; 91US-0792473.

XX PA (INSP) INST PASTEUR.

XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX PI Kourilsky P, Mottez E, Abastado J;

XX DR WPI; 2000-037081/03.

XX PT Composition containing an antigen and altered major histocompatibility
XX PT Class II determinant, used to immunize against autoimmune diseases,
XX PT e.g. acquired immune deficiency syndrome -

XX PS Claim 8; Column 13; 96pp; English.

XX CC The present invention describes a composition capable of eliciting
XX CC anti-major histocompatibility (MHC) antibodies. The composition
XX CC comprises an antigen associated with an altered MHC Class II determinant
XX CC (I) comprising alpha1, alpha2, beta1 and beta2 polypeptide domains
XX CC encoded by a mammalian MHC Class II locus covalently linked to form a
XX CC polypeptide (I) containing beta2, alpha2, alpha1 and beta1 domains in
XX CC sequence. The resulting Antigen-MHC complex is recognizable by the T cell
XX CC receptor. The compositions are used for immunisation against, or
XX CC treatment of, a wide range of autoimmune diseases, e.g. acquired immune
XX CC deficiency syndrome (AIDS), lupus erythematosus, multiple sclerosis,
XX CC thyroiditis, toxic shock, tumour and snakebite, depending on the nature
XX CC of antigen. (I) is also used to analyse functional interactions between
XX CC the various domains and for targeting lymphocyte receptors. Antibodies
XX CC against (I) are produced by usual methods of immunisation or cell fusion,
XX CC and may be humanised by standard methods. These antibodies are useful for
XX CC diagnosis (detection or purification of MHC gene products), therapy
XX CC (neutralising MHC on cell surfaces) and in the study of MHC and cellular
XX CC processes. AA233240 to AA233242 and AAY52840 to AAY52970 represent
XX CC sequences used in the exemplification of the present invention.

XX SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 21; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFEKL 8
| | | | |
Db 1 SIINFEKL 8

RESULT 27

AAG77871
ID AAG77871 standard; Protein; 8 AA.

XX AC AAG77871;

XX DT 08-MAY-2002 (first entry)

XX DE Ovalbumin 257-264 peptide.

XX KW Ovalbumin 257-264 peptide; major histocompatibility complex;
XX KW MHC-peptide complex; MHC; human; MHC class I alpha chain;
XX KW beta-2 microglobulin; MHC class II alpha chain; MHC class II beta chain;
XX KW vaccine; immune response modulation; hyperproliferative disorder;
XX KW neoplasm; hypergamaglobulinaemia; viral infection; hepatitis;
XX KW meningitis; bacterial infection; tuberculosis; gingivitis;
XX KW parasitic infection; autoimmune disease; Hashimoto's disease;
XX KW Graves' disease; rheumatoid arthritis; allergy; asthma; organ rejection;
XX KW graft-versus-host disease; GVHD; antigenic peptide.

XX OS Aves.

XX PN WO200178768-A2.

XX PD 25-OCT-2001.

XX PF 12-APR-2001; 2001WO-US11912.

XX PR 12-APR-2000; 2000US-196472P.

XX PA (UTRP) UNIV ROCHESTER.

XX PI Zauderer M, Smith ES;

XX DR WPI; 2001-602927/68.

XX PT Novel compound comprising major histocompatibility complex-peptide
XX PT complexes, used to modulate immune responses -

XX PS Example 19; Page 101; 166pp; English.

XX CC The invention comprises a compound which contains one or more major
XX CC histocompatibility complex (MHC)-peptide complexes, and an antibody
XX CC specific for a cell surface marker. The complexes comprise an MHC class
XX CC I alpha chain, a beta-2 microglobulin molecule and an antigenic peptide
XX CC bound in the MHC groove. Alternatively, the complexes may comprise an MHC
XX CC class II alpha chain, an MHC class II beta chain, and an antigenic
XX CC peptide bound in the MHC groove. The complexes are linked to the carboxyl
XX CC terminus of the antibody. The compounds of the invention can be used as a
XX CC vaccine to modulate an immune response. The compounds of the invention
XX CC are useful for treating hyperproliferative disorders (e.g. neoplasms and
XX CC hypergamaglobulinaemia); viral infections (e.g. hepatitis and
XX CC meningitis); bacterial infections (e.g. tuberculosis and gingivitis);
XX CC parasitic infections; autoimmune diseases (e.g. Hashimoto's disease;
XX CC Graves' disease and rheumatoid arthritis); allergic reactions/conditions
XX CC (e.g. asthma). The compounds of the invention may also be used in the
XX CC treatment of organ rejection or graft-versus-host disease (GVHD). The
XX CC present amino acid sequence represents the ovalbumin 257-264 peptide,
XX CC which was used as an antigenic peptide in an example of the invention.

XX SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 22; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 SIINFEKL 8
Db      1 SIINFEKL 8

RESULT 28
AAM52562
ID AAM52562 standard; Peptide; 8 AA.
XX
AC AAM52562;
XX
XX
DT 04-FEB-2002 (first entry)
XX
DE Cytotoxic T-cell epitope for ovalbumin.
XX
KW Cell death; toxic gene; tumour suppressor; ovalbumin;
KW Cytotoxic T-cell epitope.
XX
OS Unidentified.
XX
XX WO200172995-A2.
XX
XX
XX PD 04-OCT-2001.
XX
XX PF 28-MAR-2001; 2001WO-US09953.
XX
XX PR 28-MAR-2000; 2000US-0192586.
XX PR 10-MAY-2000; 2000US-0203343.
XX PR 23-JAN-2001; 2001US-0263226.
XX PR 27-FEB-2001; 2001US-0271426.
XX
XX PA (UYRP ) UNIV ROCHESTER.
XX
XX PI Zauderer M, Smith ES;
XX
XX DR WPI; 2001-570897/64.
XX
XX PT Selecting target polynucleotides, particularly toxic genes, involves
XX PT introducing a library of insert polynucleotides into a host cell
XX PT population, where the target polynucleotide promotes cell death -
XX
XX PS Example 1; Page 136; 359pp; English.
XX
XX CC The present invention relates to a method for selecting a target
XX CC polynucleotide. The method comprises introducing into a host cell
XX CC a population a library of insert polynucleotides, where expression of the
XX CC target polynucleotide directly or indirectly promotes host cell death.
XX CC The cells are cultured and the insert polynucleotides are collected from
XX CC the cells which die. The method is useful for selecting target
XX CC polynucleotides, particularly polynucleotides which alter cell phenotypes
XX CC of induce or inhibit cell death. The method can be used to isolate toxic
XX CC genes such as tumour suppressors. The present sequence was used to
XX CC illustrate the method of the the present invention.
XX
XX SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIINFEKL 8
Db      1 SIINFEKL 8

RESULT 29
AAE13119
ID AAE13119 standard; peptide; 8 AA.
XX
AC AAE13119;
XX
XX DT 28-JAN-2002 (first entry)
XX

```

```

XX
DE Ovalbumin (OVA)-derived immunodominant octapeptide.
XX
KW Immunostimulatory fusion protein; IFP; antigen component; therapy;
KW immunostimulatory component; T-cell mediated immune response; DC;
KW dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
KW Ovalbumin; OVA-derived immunodominant octapeptide.
XX
OS Unidentified.
XX
XX WO200174855-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US10515.
XX
XX PR 30-MAR-2000; 2000US-193504P.
XX
XX PA (DEND-) DENDREON CORP.
XX
XX PI Laus R, Vidovic D, Graddis T;
XX
XX DR WPI; 2001-662965/76.
XX
XX PT An immunostimulatory fusion protein comprising the intracellular domain
XX PT of HER-2 and an antigen elicits an immune response to the antigen and
XX PT is useful for the treatment of associated cancer associated -
XX
XX PS Example 1; Page 25; 59pp; English.
XX
XX CC The invention relates to immunostimulatory fusion proteins (IFP) and
XX CC nucleic acid molecules encoding such proteins. The IFPs comprise a
XX CC polypeptide antigen component and an immunostimulatory component derived
XX CC from the intracellular domain of HER-2 protein which is effective to
XX CC elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
XX CC immune response to the antigen. IFP or superactivated dendritic cells
XX CC are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
XX CC associated with a particularly antigen. The present sequence is a
XX CC ovalbumin (OVA)-derived immunodominant octapeptide. This peptide
XX CC is used in the fusion constructs of the invention.
XX
XX SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIINFEKL 8
Db      1 SIINFEKL 8

RESULT 30
AAE12145
ID AAE12145 standard; peptide; 8 AA.
XX
XX AC AAE12145;
XX
XX DT 15-JAN-2002 (first entry)
XX
XX DE Murine ovalbumin (OVA) peptide.
XX
XX KW Microbial delivery vehicle; prophylactic; immunisation; gene therapy;
XX KW tumour; carcinoma; neurodegeneration; muscular atrophy; cytostatic;
XX KW neuroprotective; antibacterial; insecticide; fungicide; antiviral;
XX KW antiprotozoal; cytostatic; anti-inflammatory; murine; ovalbumin; OVA;
XX KW listeriolysin O; LLO; MHC; major histocompatibility complex.
XX
XX OS Mus sp.
XX
XX PN US6287556-B1.
XX
XX PD 11-SEP-2001.

```

XX PF 21-DEC-1999; 99US-0469197.
 XX PR 13-AUG-1998; 98US-0133914.
 XX PA (REGC) UNIV CALIFORNIA.
 XX PI Portnoy DA, Higgins DE;
 XX DR WPI; 2001-647179/74.
 XX PT Vaccine for preventing (e.g. as immunizations) or treating (e.g. as
 PT gene therapy) tumor, carcinoma, neurodegeneration or muscular atrophy,
 PT comprises a non-virulent bacterium -
 XX Example; Column 12; 14pp; English.
 XX The invention relates to microbial-based intracellular delivery of
 CC agents to eukaryotic cells. The agents include microbial delivery
 CC vehicles such as nonvirulent bacteria comprising a first gene
 CC encoding a nonsecreted foreign cytolysin operably linked to a
 CC heterologous promoter and a second gene encoding a different
 CC foreign agent. The foreign agent may be a nucleic acid or protein,
 CC and is frequently bioactive in and therapeutic to the target
 CC eukaryote. The vaccine comprising nonvirulent bacterium is useful
 CC for prophylactics (e.g. as immunisations) and treatments (e.g. as
 CC gene therapy) of e.g. tumour, carcinoma, neurodegeneration or
 CC muscular atrophy. The present sequence is murine ovalbumin (OVA)
 CC peptide. This sequence is used to examine the ability of
 CC Escherichia coli expressing listeriolysin O (LLO) and an antigenic
 CC protein to deliver the antigen to the cytosol of macrophages for
 CC processing and presentation on MHC (major histocompatibility
 CC complex) class I molecules.
 XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 SIINFEKL 8
 Db |||||
 1 SIINFEKL 8
 RESULT 31
 AAE09514 standard; peptide; 8 AA.
 XX AC AAE09514;
 XX DT 19-NOV-2001 (first entry)
 XX DE Human ovalbumin peptide.
 XX KW Mucin; cytostatic; immunostimulant; cell mediated immune response;
 KW carcinoma; adenocarcinoma; breast cancer; dendritic cell; vaccine;
 KW gene therapy; human; ovalbumin.
 XX OS Homo sapiens.
 XX FN WO200157068-A1.
 XX PD 09-AUG-2001.
 XX PF 01-FEB-2001; 2001WO-AU00090.
 XX PR 01-FEB-2000; 2000AU-0005369.
 XX PR 14-JUN-2000; 2000US-0593870.
 XX PA (AUST-) AUSTIN RES INST.
 XX PI McKenzie IFC, Pietersz GA, Apostolopoulos V;

XX WPI; 2001-541537/60.
 XX Immunostimulant peptide, used as an anti-carcinoma vaccine, comprises a
 PT an epitope of the non-VNTR, non-leader region of a mucin -
 XX Disclosure; Page 14; 84pp; English.
 XX The patent discloses peptide or polypeptides capable of eliciting
 CC an immune response, comprising an amino acid sequence corresponding
 CC to an epitope of the non-central portion of varying numbers of an
 CC amino acid motif (VNTR), non-leader region of a mucin. The peptides
 CC of the invention, fusion proteins comprising the peptide and conjugate
 CC compounds with carbohydrate polymers are used to induce a cell mediated
 CC immune response against mucin in the prevention or treatment of
 CC carcinoma, preferably adenocarcinoma, most preferably breast cancer.
 CC They are also used to pulse dendritic cell for in vivo transfer and
 CC use as a vaccine. They are also used in gene therapy. The present
 CC sequence is ovalbumin peptide from human. This sequence is used for
 CC the prediction of T-cell epitopes.
 XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 SIINFEKL 8
 Db |||||
 1 SIINFEKL 8
 RESULT 32
 AAG63855 standard; peptide; 8 AA.
 XX AC AAG63855;
 XX DT 29-OCT-2001 (first entry)
 XX DE Amino acid sequence of an OVA-derived minimal CTL peptide.
 XX KW Opi; lipoprotein; adjuvant; type1 immune response; gp63;
 KW Leishmania major; leishmaniasis; TBC; leprosy; mycotin infection;
 KW allergic asthma; autoimmune disease.
 XX OS Synthetic.
 XX FN WO200160404-A2.
 XX PD 23-AUG-2001.
 XX PF 13-FEB-2001; 2001WO-EP01673.
 XX PR 18-FEB-2000; 2000EP-0200589.
 XX PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
 XX PI Revets H, Cornelis P, De Baetselier P;
 XX WPI; 2001-522552/57.
 XX Use of major Opi lipoprotein of Pseudomonas aeruginosa or its
 PT functional fragments as adjuvant to obtain a Th1 type immune response
 PT against heterologous antigen, for treating leishmaniasis, leprosy,
 PT allergic asthma -
 XX Disclosure; Page 12; 54pp; English.
 XX The present sequence represents an OVA-derived minimal CTL peptide,
 CC which was used in the course of the invention. The specification
 CC describes the use of the major Opi lipoprotein of Pseudomonas
 CC aeruginosa or its functional fragments as an adjuvant to obtain

```

CC a Th1 type immune response against a heterologous antigen. They are
CC especially used as an adjuvant to obtain a Th1 type immune response
CC against a heterologous antigen such as antigen gp63 of Leishmania major,
CC for treating a disease such as leishmaniasis, TBC (undefined), leprosy,
CC mycotin infection, allergic asthma or an autoimmune disease, in which
CC the natural Th1 response is insufficient and/or in which the immune
CC response is polarizes towards Th2 response.
XX
SQ Sequence 8 AA;
  Query Match 100.0%; Score 38; DB 22; Length 8;
  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 1 SIINFEKL 8
  DB 1 SIINFEKL 8

RESULT 33
AAG66422
ID AAG66422 standard; Peptide; 8 AA.
XX
AC AAG66422;
XX
DT 23-OCT-2001 (first entry)
XX
DE Chicken ovalbumin peptide, OVA257-264, used as a peptide antigen.
XX
KW Immunomodulator; vaccine; immune response; immunogenic; chicken;
KW ovalbumin.
XX
OS Gallus domesticus.
XX
PN WO200154720-A1.
XX
PD 02-AUG-2001.
XX
PF 05-JAN-2001; 2001WO-EP00087.
XX
PR 28-JAN-2000; 2000AT-0000129.
XX
PA (CIST-) CISTEM BIOLOGICALS GMBH.
XX
PI Lingnau K, Matner F, Schmidt W, Birnstiel M, Buschle M;
XX
DR WPI; 2001-536419/59.
XX
PT Pharmaceutical composition useful for inducing immune response
PT comprises antigen, immunogenic oligodeoxynucleotide containing
PT cytosine-guanine dinucleotide motifs and polycationic polymer -
XX
PS Example 1; Page 22; 39pp; English.
XX
CC The present invention relates to a pharmaceutical composition which
CC comprises an antigen, an immunogenic oligodeoxynucleotide containing
CC cytosine-guanine dinucleotide (CpG) motifs (CpG-ODN) and a polycationic
CC polymer. The composition is useful for making a vaccine to induce potent
CC immune responses, or to decrease or ablate undesired immune responses.
CC This present sequence, OVA257-264, is a peptide from chicken ovalbumin.
CC This sequence was used as a peptide antigen in the method of the present
CC invention.
XX
SQ Sequence 8 AA;
  Query Match 100.0%; Score 38; DB 22; Length 8;
  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 1 SIINFEKL 8
  DB 1 SIINFEKL 8

CC a Th1 type immune response against a heterologous antigen. They are
CC especially used as an adjuvant to obtain a Th1 type immune response
CC against a heterologous antigen such as antigen gp63 of Leishmania major,
CC for treating a disease such as leishmaniasis, TBC (undefined), leprosy,
CC mycotin infection, allergic asthma or an autoimmune disease, in which
CC the natural Th1 response is insufficient and/or in which the immune
CC response is polarizes towards Th2 response.
XX
SQ Sequence 8 AA;
  Query Match 100.0%; Score 38; DB 22; Length 8;
  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 1 SIINFEKL 8
  DB 1 SIINFEKL 8

RESULT 34
AAU05398
ID AAU05398 standard; peptide; 8 AA.
XX
AC AAU05398;
XX
DT 24-OCT-2001 (first entry)
XX
DE Peptide released from ovalbumin (Ova) after cleavage of peptide P1.
XX
KW Heat shock protein; hsp; CD8+ cytotoxic T lymphocyte; ovalbumin;
KW CTL; CD4+ T cell; AIDS; acquired immunodeficiency syndrome; murine;
KW human immunodeficiency virus; HIV; pathogen; cancer.
XX
OS Mus sp.
XX
PN WO200151081-A1.
XX
PD 19-JUL-2001.
XX
PF 01-DEC-2000; 2000WO-US32831.
XX
PR 14-JAN-2000; 2000US-0176143.
XX
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Huang Q, Richmond JFL, Cho BK, Palliser D, Chen J, Eisen HN;
PI Young RA;
XX
DR WPI; 2001-451815/48.
XX
PT Inducing a CD8+ cytotoxic T lymphocyte immune response in an individual
PT for treating diseases such as HIV involves administering a fusion
PT molecule comprising a heat shock protein -
XX
PS Example 1; Fig 1C; 58pp; English.
XX
CC The present sequence represents a naturally occurring murine peptide
CC which is released from ovalbumin (Ova) upon cleavage of peptide P1
CC (AAU05397). The present sequence is described in an invention
CC relating to a novel method of inducing a CD8+ cytotoxic T lymphocyte
CC (CTL) response to a molecule in an individual by administering a
CC fusion molecule joined to a hsp, or an adenosine triphosphate (ATP)
CC binding domain of a hsp. The method is particularly useful in inducing
CC a CD8+ CTL response in an individual deficient in CD4+ T cells e.g. for
CC treating an AIDS acquired immunodeficiency syndrome patient carrying
CC the human immunodeficiency virus (HIV). The method is also useful for
CC treating diseases that are caused by or associated with intracellular
CC pathogens, and for treating cancer.
XX
SQ Sequence 8 AA;
  Query Match 100.0%; Score 38; DB 22; Length 8;
  Best Local Similarity 100.0%; Pred. No. 9.3e+05;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 1 SIINFEKL 8
  DB 1 SIINFEKL 8

RESULT 35
AAE06033
ID AAE06033 standard; peptide; 8 AA.
XX
AC AAE06033;
XX
DT 25-SEP-2001 (first entry)
XX
DE Chicken ovalbumin CTL epitope.
XX

```


PT controlling immune responses -
 XX Example 6; Page 69; 97pp; English.
 XX
 CC The present invention describes immunomodulatory populations (I) and
 CC (II) of conjugate molecules (Cm) comprising immunostimulatory sequences
 CC (ISS) of polynucleotides and antigens. The extent of conjugation affects
 CC the immunological properties (e.g. the extent of antigen-specific
 CC antibody formation, including Th1-associated antibody formation) so the
 CC conjugates are used for altering the type and extent of immune response.
 CC (I) and (II) have immunomodulatory, immunosuppressive and anti-allergic
 CC activities, and can be used in the modulation of immune responses via
 CC the stimulation of Th1 lymphocytes and Th1-associated cytokines, and
 CC suppression of Th2 lymphocytes and cytokines. The populations (I) and
 CC (II) of conjugate molecules may be used for modulating immune responses
 CC in individuals e.g. for the treatment of an allergic condition. (I) and
 CC (II) may be used to modulate immune responses and therefore prevent
 CC potentially harmful reactions to antigens. The present sequence
 CC represents an ovalbumin (OVA) cytotoxic T lymphocyte (CTL) epitope
 CC which is used in the exemplification of the present invention.
 XX
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB |||||
 1 SIINFEKL 8
 RESULT 38
 AAB82176
 ID AAB82176 standard; peptide; 8 AA.
 XX
 AC AAB82176;
 XX
 DT 04-JUL-2001 (first entry)
 XX
 DE Chicken ovalbumin (OVA) peptide.
 XX
 KW Ovalbumin; chicken; OVA; immune response; Langerhans cell migration;
 KW tumour; EG7-OVA.
 XX
 OS Gallus gallus.
 XX
 PN US6210672-B1.
 XX
 PD 03-APR-2001.
 XX
 PP 20-OCT-1998; 98US-0176044.
 XX
 PR 20-OCT-1998; 98US-0176044.
 XX
 PA (TORR-) TORREY PINES INST MOLECULAR STUDIES.
 XX
 PI Cowing C;
 XX
 XX WPI; 2001-280845/29.
 XX
 PT Enhancing an immune response to an antigen in a mammal comprises
 PT topically administering the antigen, a penetration enhancer and an
 PT agent for enhancing Langerhans cell migration -
 XX
 PS Disclosure; Column 4; 15pp; English.
 XX
 CC This invention relates to a method for enhancing an immune response to an
 CC antigen in a mammal. The method comprises administering a composition
 CC comprising the antigen, a penetration enhancer selected from lipophilic
 CC solvents, low-frequency ultrasound, electroporation, iontophoresis and
 CC intraepidermal delivery and an agent for enhancing Langerhans cell
 CC migration to an epidermal or mucosal site. The method can be used to
 CC enhance the immune response to tumours, viruses, bacteria and parasites.
 CC The present sequence represents a fragment of the chicken ovalbumin (OVA)
 CC protein. The peptide functions as a fragment of the chicken ovalbumin (OVA)
 CC antigen for CD8+ cytotoxic T lymphocytes. The peptide can be used in the
 CC method of the invention to enhance an immune response to the EG7-OVA
 CC tumour.
 XX
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB |||||
 1 SIINFEKL 8

CC is useful for conferring protective immunity against an intracellular
 CC pathogen. The present peptide: immunodominant CTL epitope of ovalbumin,
 CC was used to illustrate the present invention. This peptide was used to
 CC test for the ability of archaeosomes to induce CTL responses to
 CC ovalbumin.
 XX
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB |||||
 1 SIINFEKL 8
 RESULT 39
 AAB81122
 ID AAB81122 standard; peptide; 8 AA.
 XX
 AC AAB81122;
 XX
 DT 04-JUL-2001 (first entry)
 XX
 DE Chicken ovalbumin (OVA) peptide.
 XX
 KW Ovalbumin; chicken; OVA; immune response; Langerhans cell migration;
 KW tumour; EG7-OVA.
 XX
 OS Gallus gallus.
 XX
 PN US6210672-B1.
 XX
 PD 03-APR-2001.
 XX
 PP 20-OCT-1998; 98US-0176044.
 XX
 PR 20-OCT-1998; 98US-0176044.
 XX
 PA (TORR-) TORREY PINES INST MOLECULAR STUDIES.
 XX
 PI Cowing C;
 XX
 XX WPI; 2001-280845/29.
 XX
 PT Enhancing an immune response to an antigen in a mammal comprises
 PT topically administering the antigen, a penetration enhancer and an
 PT agent for enhancing Langerhans cell migration -
 XX
 PS Disclosure; Column 4; 15pp; English.
 XX
 CC This invention relates to a method for enhancing an immune response to an
 CC antigen in a mammal. The method comprises administering a composition
 CC comprising the antigen, a penetration enhancer selected from lipophilic
 CC solvents, low-frequency ultrasound, electroporation, iontophoresis and
 CC intraepidermal delivery and an agent for enhancing Langerhans cell
 CC migration to an epidermal or mucosal site. The method can be used to
 CC enhance the immune response to tumours, viruses, bacteria and parasites.
 CC The present sequence represents a fragment of the chicken ovalbumin (OVA)
 CC protein. The peptide functions as a fragment of the chicken ovalbumin (OVA)
 CC antigen for CD8+ cytotoxic T lymphocytes. The peptide can be used in the
 CC method of the invention to enhance an immune response to the EG7-OVA
 CC tumour.
 XX
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB |||||
 1 SIINFEKL 8

CC The present invention relates to a vaccine composition comprising a
 CC liposome prepared from the total polar lipids extract of an
 CC archaeobacterium and an acellular antigen, preferably an isolated
 CC outer membrane from a pathogen. The vaccine of the invention provides an
 CC enhanced cytotoxic T lymphocyte response. The vaccine of the invention

Db 1 SIINFEKL 8

RESULT 40
AAB82065
ID AAB82065 standard; peptide; 8 AA.
AC AAB82065;
DT 22-JUN-2001 (first entry)
XX Ovalbumin-derived peptide, used as a control peptide.
DE Antigen; immunostimulant; vaccine; pharmaceutical composition; antiviral;
KW viral infection; ovalbumin.
XX Unidentified.
OS WO200124822-A2.
XX 12-APR-2001.
XX 02-OCT-2000; 2000WO-EP09657.
XX 01-OCT-1999; 99AT-0001680.
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
PI Fleitmann J, Mattner F, Buschle M, Melling J;
XX WPI; 2001-290577/30.
XX New pharmaceutical composition comprising an antigen, an
PT immunostimulating substance and a polycationic polymer, useful in
PT manufacturing vaccines -
XX Example 1; Page 11; 20pp; English.
XX The present invention relates to a pharmaceutical composition comprising
CC (a) an antigen; (b) an immunostimulating substance consisting of
CC neuroactive compounds, hormones, compounds having growth hormone activity
CC or their mixtures; and (c) a polycationic polymer. The composition is
CC useful in manufacturing vaccines. To illustrate the present invention, a
CC murine tyrosinase related protein-2 peptide (TRP-2 peptide; see
CC AAB82064) was used. Mice were injected subcutaneously with either the
CC TRP-2 peptide, TRP-2 peptide + human growth hormone (hGH), TRP-2 peptide
CC + poly-L-arginine 60 (pR60) or TRP-2 peptide + pR60 + hGH. Animals were
CC sacrificed 10 days post injection, and mesenteric and inguinal lymph
CC nodes were harvested. Lymphocytes were prepared from lymph nodes and were
CC re-stimulated with TRP-2 peptide or with an ovalbumin-derived peptide
CC (the present peptide), with the same major histocompatibility complex
CC (MHC) restriction serving as negative control. Spots representing single
CC T cells specific for the peptide used for re-stimulation were counted. No
CC spots were detected when the ovalbumin derived peptide was used, while
CC TRP-2 peptide + pR60 + hGH showed the highest number of spots or single T
CC cells. The present peptide was also used as a control peptide for
CC experiments with substance P (see AAB82070).
XX Sequence 8 AA;
Query Match 100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 1 SIINFEKL 8
RESULT 41
AAB92374
ID AAB92374 standard; Peptide; 8 AA.
XX

AC AAB92374;
XX 22-JUN-2001 (first entry)
XX Miscellaneous peptide SEQ ID NO:1550.
DE Protection; endogenous therapeutic peptide; peptidase; conjugation;
KW blood component; modification; succinimide; maleimido group; amino;
KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.
XX Homo sapiens.
OS Synthetic.
XX WO2000069900-A2.
XX 23-NOV-2000.
XX 17-MAY-2000; 2000WO-US13576.
XX 17-MAY-1999; 99US-0134406.
XX 10-SEP-1999; 99US-0153406.
XX 15-OCT-1999; 99US-0159783.
XX (CONJ-) CONJUCHEM INC.
XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thiabaudau K;
PI WPI; 2001-112059/12.
XX Modifying and attaching therapeutic peptides to albumin prevents
PT peptidase degradation, useful for increasing length of in vivo activity
PT
XX Disclosure; Page 711; 733pp; English.
XX The present invention describes a modified therapeutic peptide (I)
CC comprising a therapeutically active amino acid region (III) and a
CC reactive group (II) (e.g. succinimide and maleimido groups) attached to
CC bonds with amino/hydroxyl/thiol groups on blood components to form a
CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
CC factors and neurotransmitters, to protect them from peptidase activity
CC in vivo for the treatment of various disorders. Endogenous therapeutic
CC peptides are not suitable as drug candidates as they require frequent
CC administration due to rapid degradation by peptidases in the body.
CC Modifying and attaching therapeutic peptides to albumin prevents or
CC reduces the action of peptidases to increase length of activity (half
CC life) and specificity as bonding to large molecules decreases
CC intracellular uptake and interference with physiological processes.
CC AAB90829 to AAB92441 represent peptides which can be used in the
CC exemplification of the present invention.
XX Sequence 8 AA;
Query Match 100.0%; Score 38; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 1 SIINFEKL 8
RESULT 42
AAB48950
ID AAB48950 standard; Protein; 8 AA.
XX AAB48950;
AC AAB48950;
XX 27-MAR-2001 (first entry)
DT Ovalbumin MHC class I epitope, SEQ ID NO:6.
DE

XX Transposable element; MHC epitope; major histocompatibility complex;
 KW intracellular bacterial pathogen; loxp site; Cre recombinase;
 KW insertion end; in-frame fusion; detection; antigen;
 KW disseminated insertions of class-I epitopes; DICE-I; transposon Tn5;
 KW ovalbumin MHC class I epitope.
 XX Unidentified.
 OS WO200071158-A1.
 XX 30-NOV-2000.
 XX 26-MAY-2000; 2000WO-US14687.
 PF 26-MAY-1999; 99US-0136210.
 PR (UYOR-) UNIV OREGON HEALTH SCI.
 PA Heffron FL, Parker DC, Ellefson DD;
 PI WPI; 2001-031967/04.
 DR Transposable element for detecting an antigenic epitope of a pathogen,
 PT comprising 5' and 3' recombining sites, nucleic acid sequences encoding
 PT a selectable marker and major histocompatibility complex (MHC) epitope,
 PT and an insertion end -
 XX Claim 5; Page 41; 63pp; English.
 PS The invention relates to a novel transposable element comprising DNA
 CC encoding a selectable marker (e.g., antibiotic resistance) located
 CC between a 5' recombining site and a 3' recombining site (e.g., loxp
 CC sites); DNA encoding an MHC (major histocompatibility complex) epitope
 CC either 5' of the 5' recombining site or 3' of the 3' recombining site;
 CC and insertion ends comprising an inverted repeat sequence at the 5' and
 CC 3' ends of the transposable element sufficient for integration of the
 CC transposable element. The transposable elements of the invention are able
 CC to introduce in-frame insertions throughout the chromosome of an
 CC intracellular bacterial pathogen. This system "tags" the bacterial gene
 CC and resulting protein, allowing the identification of proteins
 CC secreted across the membranes of the eukaryotic cell infected by the
 CC bacterium. In one embodiment, the transposable elements contain an
 CC antibiotic resistance cassette, two minimal loxp recombination sites, an
 CC MHC class I or class II epitope, and flanking insertion ends. A
 CC transposase, such as the Cre recombinase protein, is expressed in trans
 CC from a plasmid, or can be included in the transposable element. The Cre
 CC recombinase loops out the intervening sequences containing the antibiotic
 CC resistance cassette. When the transposable element inserts within a gene,
 CC the resolved insertion places the MHC class I or class II epitope in
 CC frame with the gene. The transposable elements of the invention are
 CC useful for detecting an antigenic epitope of an intracellular bacterial
 CC pathogen, such as *Salmonella* sp., *Mycobacterium tuberculosis* and *Listeria*
 CC monocytogenes. Certain embodiments of the technology, termed
 CC "disseminated insertions of class-I epitopes" (DICE-I; DICE-II for
 CC class II epitopes) allow the rapid and accurate identification of
 CC proteins involved in bacterial pathogenesis so that such proteins can
 CC be used as vaccine and drug targets. Carrier vaccines may be generated
 CC by infecting bacteria with a transposable element of the invention
 CC which additionally comprises an antigen associated with a disease,
 CC preferably cancer or a viral or bacterial disease, operably linked to the
 CC MHC epitope DNA of the transposable element. The present sequence
 CC represents an ovalbumin MHC class I epitope specifically claimed
 CC for use in the invention.
 XX SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 22; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 |||||

Db 1 SIINFEKL 8
 RESULT 43
 AAE28959
 ID AAE28959 standard; peptide; 8 AA.
 XX AAE28959;
 AC 27-JAN-2003 (first entry)
 XX Chicken ovalbumin peptide.
 DE Modified vaccinia Ankara virus; MVA; HIV; human immunodeficiency virus;
 KW CD8+ T cell; immune response; acquired immune deficiency syndrome; AIDS;
 KW viral infection; vaccine; immunostimulant; virucide; chicken.
 XX Gallus sp.
 OS WO200272754-A2.
 XX 19-SEP-2002.
 XX 01-MAR-2002; 2002WO-US06713.
 PF 08-MAR-2001; 2001US-274434P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Moss B, Wyatt L, Earl P;
 PI WPI; 2002-723330/78.
 DR New recombinant modified vaccinia Ankara (MVA) virus expressing HIV
 XX env, gag and pol genes, useful for boosting or inducing CD8 T cell
 PT immune responses in primates, e.g. humans, particularly for preventing
 PT AIDS or other viral infections -
 XX Example 1; Page 111; 112pp; English.
 PS The present invention relates to a composition comprising recombinant
 CC modified vaccinia Ankara (MVA) virus expressing an HIV (human immuno-
 CC deficiency virus) env, gag and pol gene or its modified gene for the
 CC production of an HIV Env, Gag and Pol antigen by expression from the
 CC recombinant MVA virus. The HIV env gene is modified to encode an HIV
 CC Env protein composed of gp120 and the membrane-spanning and ectodomain
 CC of gp41, but lacking part or all of the cytoplasmic domain of gp41. The
 CC composition or recombinant MVA virus is useful for boosting or inducing
 CC CD8+ T cell immune response in primates, particularly in humans. The
 CC composition may be used for preventing AIDS (acquired immune deficiency
 CC syndrome) or other viral infections. Sequences of the invention are also
 CC used as vaccines. The present sequence is chicken ovalbumin peptide. This
 CC sequence is used in the exemplification of the invention.
 XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 |||||
 Db 1 SIINFEKL 8
 RESULT 44
 AAE26368
 ID AAE26368 standard; peptide; 8 AA.
 XX AAE26368;
 AC 13-DEC-2002 (first entry)
 XX

DE Ovalbumin CTL epitope.
 XX Human; immune response; T-helper cell epitope; chitosan; CTL response;
 KW vaccine; prostate cancer; breast cancer; cytostatic; immunostimulant.
 XX Unidentified.
 OS WO200234287-A2.
 PN
 XX
 XX 02-MAY-2002.
 PD
 XX 26-OCT-2001; 2001WO-DK00705.
 PF
 XX 27-OCT-2000; 2000DK-0001606.
 XX 03-NOV-2000; 2000US-245166P.
 PR 18-JUN-2001; 2001DK-0000936.
 PR
 XX (PHAR-) PHARMEXA AS.
 XX
 XX Beier AM, Gautam A, Mouritsen S;
 PI
 XX WPI; 2002-463339/49.
 XX
 XX Inducing or enhancing an immune response against an antigen,
 DR particularly cytotoxic T-lymphocyte responses, for treating or
 XX ameliorating prostate or breast cancer, comprises administering the
 XX antigen formulated with chitosan -
 XX
 XX Example 3; Page 63; 97pp; English.
 PS
 XX The invention relates to a method for inducing or enhancing an immune
 CC response against a polypeptide antigen in an animal, including human.
 CC The method comprises administering the polypeptide antigen or at least
 CC one variant which includes at least one first T-helper cell epitope that
 CC is foreign to the animal (foreign TH epitope) and is formulated with
 CC chitosan. The polypeptide antigen is weakly immunogenic or non-
 CC immunogenic. The invention is used as vaccine. The chitosan and
 CC polypeptide antigen or its variant are useful in the preparation of an
 CC immunogenic composition for inducing or enhancing an immune response,
 CC particularly CTL response, against the polypeptide or protein antigen.
 CC The method for inducing or enhancing an immune response is useful in
 CC treating or ameliorating cancer, e.g. prostate or breast cancer. The
 CC present sequence is ovalbumin CTL epitope used to illustrate the method
 CC of the invention.
 XX
 XX Sequence 8 AA;
 SQ
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 Db |||||
 1 SIINFEKL 8
 RESULT 45
 ABB79933
 ID ABB79933 standard; Peptide; 8 AA.
 XX
 XX ABB79933;
 AC
 XX 12-DEC-2002 (first entry)
 DT
 XX Ovalbumin T-cell epitope.
 DE
 XX Vaccine; genetic immunisation; gene therapy; antigen; epitope;
 KW T-cell; T-lymphocyte; ovalbumin.
 KW Unidentified.
 OS
 XX US2002115625-A1.
 PN
 XX

PD 22-AUG-2002.
 XX
 XX 08-MAR-2001; 2001US-0801540.
 PF
 XX 19-MAY-1999; 99US-0308511.
 PR
 XX (BOTA/) BOT A.
 PA (BONA/) BONA C.
 PA
 XX Bot A, Bona C;
 PI
 XX WPI; 2002-712482/77.
 DR
 XX Immunizing an infant mammal against a target antigen or inducing a
 XX cytotoxic T cell response against a pathogen in the mammal, comprises
 PT administering nucleic acid encoding relevant epitopes of pathogen
 PT associated target antigens -
 PT
 XX Disclosure; Page 4; 45pp; English.
 PS
 XX The present sequence is that of an ovalbumin T-cell epitope (amino
 CC acids 257-264). This is a T-CTL epitope which, in the context of
 CC MHC Class II self antigens, may be recognised by a cytotoxic T-cell
 CC and thereby promote CTL-mediated lysis of cells comprising the target
 CC antigen. It is an example of T-cell epitopes which may be used
 CC according to the invention. The invention relates to immunising an
 CC infant mammal against a target antigen or inducing a cytotoxic
 CC T-cell response against a pathogen. The method involves inoculating
 CC the infant with a nucleic acid encoding one or more relevant epitopes
 CC of one or more target antigens associated with the pathogen in a
 CC carrier, so that the relevant epitope(s) is expressed in the infant
 CC mammal. B- or T-cell epitopes may be used, and the pathogen may be
 CC a virus, bacterium, protozoan, fungus, yeast, or parasite. The method
 CC may reduce the need for subsequent boost administrations and may
 CC prevent the side-effects associated with live attenuated vaccines.
 CC Administration of multiple epitopes directed to antigens
 CC associated with more than one pathogen may provide an infant with a
 CC broader spectrum of protection, and may be a means for inducing an
 CC immune response to a variety of childhood pathogens.
 XX
 XX Sequence 8 AA;
 SQ
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 Db |||||
 1 SIINFEKL 8
 RESULT 46
 ABB93028
 ID ABB93028 standard; Peptide; 8 AA.
 XX
 XX ABB93028;
 AC
 XX 20-NOV-2002 (first entry)
 DT
 XX Mouse class 1 MHC molecule Kb binding ovalbumin epitope.
 DE
 XX Regulator; transcription; cell death; phenotype; molecular scaffold;
 KW gene therapy; cancer; cardiovascular disease; arrhythmia; heart failure;
 KW ischaemia; obesity; neurodegenerative disease; Alzheimer's disease;
 KW bone pathology; dermatologic disease; psoriasis; infection; AIDS;
 KW acquired immunodeficiency syndrome; cosmetic; wound healing;
 KW antibiotic transport; drug toxicity; drug resistance; immunobiology;
 KW inflammation; allergic response; human immunodeficiency virus.
 KW Unidentified.
 OS
 XX WO200262822-A2.
 PN
 XX

PD 15-AUG-2002.
 XX
 XX
 PF 04-FEB-2002; 2002WO-US02814.
 XX
 PR 02-FEB-2001; 2001US-265589P.
 PR 05-FEB-2001; 2001US-265880P.
 PR 27-FEB-2001; 2001US-271423P.
 PR 23-JAN-2001; 2001US-263226P.
 PR 28-MAR-2000; 2000US-192586P.
 PR 22-SEP-1997; 97US-935377P.
 XX
 XX (UYRP) UNIV ROCHESTER.
 XX
 XX Zauderer M, Smith ES;
 XX
 DR WPI; 2002-643398/69.
 XX
 XX Identifying regulator polypeptides which influence target
 PT transcriptional regulatory regions, useful for treating cancer,
 PT comprises introducing host cells expressing the polypeptide into a
 PT library of polynucleotides
 XX
 XX Example 1; Page 82; 224pp; English.
 XX
 CC The invention discloses a method for identifying polynucleotides encoding
 CC a regulator polypeptide, whose expression induces activation of a target
 CC transcriptional regulatory region in a host cell. The method comprises
 CC providing a population of eukaryotic host cells capable of expressing the
 CC polypeptide, introducing into the host cell a library of polynucleotides
 CC encoding the polypeptides, permitting expression of the polypeptides and
 CC then recovering them from the host cells. The target transcriptional
 CC regulatory region is operably associated with a polynucleotide encoding a
 CC gene product, the expression of which results in host cell death or cause
 CC the host cells to exhibit a pre-determined modified phenotype and where
 CC the gene product is expressed upon activation of target transcriptional
 CC regulatory region. Each candidate regulator polypeptide comprises a
 CC candidate peptide and a molecular scaffold fused to the peptide so that
 CC the peptide is displayed on the surface of the candidate regulator
 CC polypeptide. The methods are useful in selecting and/or screening
 CC regulator molecules, such as polypeptides, which directly or indirectly
 CC induce or suppress the transcriptional activation of a target
 CC transcriptional regulatory region in a eukaryotic host cell. These
 CC regulator molecules may be used (e.g. in gene therapy) for preventing or
 CC treating cancers (e.g. breast or ovarian cancer), cardiovascular diseases
 CC (e.g. arrhythmia, heart failure, ischaemia), obesity, neurodegenerative
 CC diseases (e.g. Alzheimer's disease), bone pathologies, dermatologic
 CC diseases (e.g. psoriasis), infections (e.g. viral, bacterial), acquired
 CC immunodeficiency syndrome (AIDS), in cosmetic applications and in wound
 CC healing. The method is also useful in screening regulator molecules that
 CC block antibiotic transport mechanisms, in drug toxicities and drug
 CC resistance applications and in improving the performance of existing or
 CC developmental drugs. It may also be used in immunobiology, inflammation,
 CC allergic response and in biotechnology applications. The sequences
 CC presented in ABG92946-ABG93029 are examples of regulator polypeptides.
 XX
 XX Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. NO. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 |||||
 Db 1 SIINFEKL 8
 RESULT 47
 AAE25400
 ID AAE25400 standard; peptide; 8 AA.
 AC AAE25400;
 XX
 XX 30-OCT-2002 (first entry)
 XX

XX Ovalbumin peptide used in the invention.
 DE
 XX
 KW Recombinant vector; coat protein; CP; viral replication; infection;
 KW Zucchini yellow mosaic potyvirus; ZYMV; cucurbit fruit; vaccination;
 KW pharmaceutical; diagnostic; ovalbumin.
 XX
 OS Unidentified.
 XX
 PN WO200244323-A2.
 XX
 XX 06-JUN-2002.
 XX
 XX 28-NOV-2001; 2001WO-JL01098.
 XX
 PR 28-NOV-2000; 2000US-253136P.
 PR 27-SEP-2001; 2001US-0963761.
 XX
 XX (VIRO-) VIROGENE LTD.
 XX
 PI Gal-On A, Shibolet Y, Arazi T, Ilan Y;
 XX
 XX WPI; 2002-537446/57.
 DR N-PSDB; AAD41429.
 XX
 XX Novel recombinant vector useful for transiently expressing heterologous
 PT peptide in plant comprises potyvirus nucleic acid sequence and
 PT heterologous sequence inserted at amino terminus of potyvirus coat
 PT protein -
 XX
 XX Claim 17; Page 60; 61pp; English.
 XX
 CC The invention relates to a recombinant vector for expressing a
 CC heterologous peptide at the amino-terminus of a potyvirus coat protein
 CC (CP). The vector includes sufficient potyvirus nucleic acid sequence
 CC to permit viral replication and spread within a plant infected by the
 CC vector. The invention also relates to Zucchini yellow mosaic potyvirus
 CC (ZYMV) AGII strain CP and its corresponding nucleic acid sequence. The
 CC recombinant vector is useful for transiently expressing a portion of
 CC the heterologous peptide in a plant. It is also useful for infecting a
 CC cucurbit fruit, is useful as a source of material for vaccination,
 CC pharmaceutical or diagnostic application. The present sequence is a
 CC ovalbumin peptide used to fuse to the N-terminus of ZYMV AGII
 CC strain CP.
 XX
 XX Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. NO. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 |||||
 Db 1 SIINFEKL 8
 RESULT 48
 ABG31661
 ID ABG31661 standard; Peptide; 8 AA.
 XX
 AC ABG31661;
 XX
 XX 05-NOV-2002 (first entry)
 XX
 XX Chicken ovalbumin MHC class I restricted epitope.
 DE
 XX
 KW Chicken; ovalbumin; MHC class I; major histocompatibility complex;
 KW polycationic compound; allergen; cytokine; chemokine; wound healing;
 KW cytotoxic drug; anti-oligogenic drug; immunostimulant; antiallergic;
 KW cytostatic; vulnerary; immunogenic.
 XX
 XX Gallus gallus.
 OS
 XX

PT New isolated mutant major histocompatibility complex class I chimeric
PT protein displayed on surfaces of recombinant yeast cells, has improved
PT stability, and is useful for activating immune response -
XX
XX
PS Example 3; Figure 18; 96pp; English.
XX
CC The present invention relates to a new mutant major histocompatibility
CC complex (MHC) class I chimeric protein. The protein of the invention
CC comprises a portion mediating binding to surfaces of recombinant yeast
CC cells and a portion comprising peptide binding region of MHC class I
CC protein, where the invention is improved in stability as compared with
CC MHC class I chimeric protein which is not a mutant chimeric protein.
CC The protein, further comprising a detectable label, is useful for
CC detecting a lymphocyte having a T-cell receptor protein in a biological
CC sample such as cells, tissue sample, biopsy material or bodily fluids.
CC The method is useful for detecting a T lymphocyte that is specific for
CC a neoplastic cell, a tumour cell, a virus-infected cell, a protozoan-
CC infected cell, a bacterium-infected cell or a fungus-infected cell. The
CC protein of the invention can be used to directly activate T cells, in
CC order to identify/screen for peptide-MHC antigens. The protein is also
CC useful in activating T cells that participate in the removal of target
CC cells including neoplastic cells and cells infected with pathogenic
CC agents including viruses, protozoans, bacteria, fungi or nematodes.
CC The invention is improved in stability as compared with MHC class I
CC protein which is not a mutant chimeric protein. The present amino acid
CC sequence represents a mouse MHC peptide of the invention, as described
CC above.
XX
XX
SQ Sequence 8 AA;
Query Match 100.0%; Score 38; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
Db 1 SIINFEKL 8
RESULT 51
ABB08108
ID ABB08108 standard; peptide; 8 AA.
XX
AC ABB08108;
XX
XX 10-SEP-2002 (first entry)
XX
XX Chicken ovalbumin CTL epitope fragment.
DE T cell; antigen; tumour; vaccine; cytostatic; cancer; ovalbumin;
KW CTL epitope.
KW
OS Gallus sp.
XX
XX US6387701-B1.
XX
XX 14-MAY-2002.
XX
XX 30-APR-1999; 99US-0302329.
XX
XX 30-APR-1996; 96US-0640444.
XX 30-APR-1997; 97WO-US07317.
PR 06-MAY-1998; 98US-0073819.
PR 16-FEB-1999; 99US-0171916.
XX
XX (UYDU-) UNIV DUKE.
PA
XX Nair SK, Boczkowski DJ, Gilboa E;
XX
XX WPI; 2002-478447/51.
DR
XX Identifying tumor antigens that elicit T cell responses and which may
PT be used for vaccinating against cancers, e.g. melanomas, breast

PT cancers, prostate cancers, colon cancers, and ovarian cancers,
PT comprises a cytotoxicity assay -
XX
XX Example 1; Column 12; 21pp; English.
PS
XX The invention relates to identifying a tumour antigen that elicits a T
CC cell response directed against the tumour. The antigen may then be used
CC to vaccinate against cancers e.g. melanomas, bladder cancers, breast
CC cancers, pancreatic cancers. The present sequence represents a chicken
CC ovalbumin CTL epitope fragment.
XX
XX
SQ Sequence 8 AA;
Query Match 100.0%; Score 38; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
Db 1 SIINFEKL 8
RESULT 52
ABB81273
ID ABB81273 standard; peptide; 8 AA.
XX
XX ABB81273;
XX
XX 20-AUG-2002 (first entry)
XX
XX Chicken OVA 257-264 peptide SEQ ID NO:1.
DE
XX Yeast; dendritic cell; vaccine; immune response; ovalbumin; antifungal;
KW immunostimulant; antibacterial; virucide; antiprotozoal; cytostatic;
KW immunisation; cell mediated immunity; infectious disease; cancer.
XX
XX Gallus gallus.
XX
XX WO200239951-A2.
XX
XX 23-MAY-2002.
XX
XX 15-NOV-2001; 2001WO-US43537.
XX
XX 15-NOV-2000; 2000US-249173P.
XX
XX (GLOB-) GLOBE IMMUNE INC.
XX (UYTE-) UNIV TECHNOLOGY CORP.
XX
XX Duke RC, Bellgrau D, Franzusoff A, Wilson CC;
XX WPI; 2002-479895/51.
DR
XX Therapeutic composition, useful as vaccine, comprises dendritic cell
PT intracellularly loaded with yeast vehicle and at least one antigen -
XX
XX Example 4; Page 34; 68pp; English.
PS
XX The present invention describes a therapeutic composition (I) comprising
CC a dendritic cell (II), a yeast vehicle (III) and at least one antigen
CC (IV), where (II) has been loaded intracellularly with (III) and (IV).
CC (I) has immunostimulant, antibacterial, antifungal, virucide,
CC antiprotozoal and cytostatic activities. (I) has many attributes that
CC make it an ideal vaccine candidate, including ease of construction, low
CC expense of mass production, biological stability and safety. No grossly
CC adverse side effects of immunisation with whole yeast were apparent at
CC the time of the initial vaccination or upon real administration. The
CC composition provides a powerful strategy for the induction of cell-
CC mediated immunity directed against a variety of infectious diseases and
CC cancer targets. The present sequence represents a chicken OVA (ovalbumin)
CC 257-264 peptide which is used in an example from the present invention.
XX
XX Sequence 8 AA;
SQ

Query Match 100.0%; Score 38; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
DB 1 SIINFEKL 8

RESULT 53
AAE22531
ID AAE22531 standard; peptide; 8 AA.

AC AAE22531;

XX 26-JUL-2002 (first entry)

DT Ovalbumin peptide to evaluate the ability of aroC/sifa (P3H8) TML mutant.

DE Sifa; attenuated microorganism; medicament; allergen; gene therapy;
KW vaccine; protein therapy; virucide; hepatotropic; antiinflammatory;
KW protozoacide; ovalbumin.

XX Unidentified.

OS WO200226251-A1.

PN 04-APR-2002.

XX 01-OCT-2001; 2001WO-GB04358.

XX 29-SEP-2000; 2000GB-0023906.

PR 14-AUG-2001; 2001GB-0019802.

XX (MICR-) MICROSCIENCE LTD.

PA Brennan FR, Dougan G;

PI WPI; 2002-339986/37.

DR An attenuated Salmonella strain, for producing an elevated immune
XX response treat diseases, comprises a mutation that disrupts expression
PT of the sifa gene, and expresses a therapeutic heterologous peptide such
PT as an antigen -

PS Example 1; Page 7; 24pp; English.

XX The present invention relates to an attenuated Salmonella microorganism
CC which comprises a mutation that disrupts the expression of the sifa gene
CC and expresses a therapeutic heterologous peptide such as an antigen. The
CC attenuated microorganisms are useful for manufacturing medicaments to
CC treat or prevent a disease which can be treated by the heterologous
CC product and to increase the MHC class I-restricted response in a patient,
CC to deliver a therapeutic polynucleotide to a host cell to treat a disease
CC which can be corrected by administering the polynucleotide and to cause
CC an increase in the MHC class I-restricted response in a patient. They
CC are also useful to deliver heterologous antigens and allergens to a
CC patient, such as hepatitis, herpes simplex and Malarial antigens. The
CC method of the invention is useful for delivery of antisense nucleotides
CC or ribozymes in gene therapy. Sequences of the invention are also used
CC as vaccines and in protein therapy. The present sequence is ovalbumin
CC peptide (OVA 257-264) used to evaluate the ability of an aroC/sifa (P3H8)
CC TML mutant to stimulate a MHC class I-restricted response to heterologous
CC antigen.

SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8

DB 1 SIINFEKL 8

RESULT 54

ABB76050

ID ABB76050 standard; Peptide; 8 AA.

XX ABB76050;

XX 12-JUL-2002 (first entry)

XX Ovalbumin, H-2kb restricted epitope.

DE Bordetella pertussis; adenylate cyclase; CyaA; adenylcyclase;
KW vector; drug delivery; antigen delivery; cell targeting; CD11b;
KW ovalbumin; epitope; chicken.

XX Gallus sp.

XX EP1189446-A1.

XX 20-MAR-2002.

XX 15-SEP-2000; 2000EP-0402562.

XX 15-SEP-2000; 2000EP-0402562.

XX (INSP) INST PASTEUR.

PA (CNRS) CENT NAT RECH SCI.

XX Leclerc C, Guernonprez P, Ladant D, Guiso N, Khelef N;

XX WPI; 2002-354020/39.

DR Use of Bordetella adenylcyclase to make proteinaceous vector, useful
XX for drug or antigen delivery, selectively targets cells that express
PT CD11b -

XX Example B; Page 14; 34pp; English.

XX The present sequence is the peptide sequence of a chicken ovalbumin,
CC H-2kb restricted epitope, which was used as an experimental model
CC epitope in an example from the invention. The epitope was
CC genetically inserted into the catalytic domain of a detoxified, but
CC still invasive, mutant adenylate cyclase (adenylcyclase, CyaA) of
CC Bordetella pertussis. The recombinant toxin, CyaAOVA, was used to
CC immunise C57BL/6(H-2b) mice once i.v. CD4- and CD40-independent
CC cytotoxic T lymphocyte (CTL) priming was observed in the absence
CC of adjuvant. The invention relates to the novel use of Bordetella
CC CyaA as a proteinaceous vector for targeting a molecule of interest
CC to the surface CD11b-expressing cells, especially dendritic cells
CC and neutrophils. The molecule of interest is translocated in the
CC cytosol to prime a CTL response. In a preferred embodiment, a
CC peptide is inserted into the catalytic domain of CyaA at a
CC permissive site. The peptide may be an intracellular bacterial
CC cell, tumour, viral, fungal or parasite cell antigen (all claimed).
CC Alternatively, a drug, especially an antiinflammatory, is chemically
CC coupled to CyaA for drug delivery.

SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8

DB 1 SIINFEKL 8

RESULT 55
AAE19945

ID AAE19945 standard; peptide; 8 AA.
 XX AAE19945;
 AC
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Cytotoxic T-cell epitope for ovalbumin.
 XX
 KW Cytotoxic T cell; CTL; tumour; cancer; infection; cell-mediated immunity;
 KW vaccine; immune response; cytostatic; T-cell epitope; ovalbumin.
 XX
 OS Unidentified.
 XX
 PN US2002018785-A1.
 XX
 PD 14-FEB-2002.
 XX
 PF 02-APR-2001; 2001US-0822250.
 XX
 PR 22-SEP-1997; 97US-0935377.
 XX
 PA (UYP) UNIV ROCHESTER.
 XX
 PI Zauderer M;
 XX
 DR WPI; 2002-239252/29.
 XX
 XX Representational Difference Analysis method for identification of
 PT antigens recognized by cytotoxic T cells and specific for human tumors,
 PT comprises improved selection of genes encoding target antigens -
 XX
 PS Example 1; Page 13; 54pp; English.
 XX
 CC The present invention relates to novel methods for the identification
 CC of antigens recognised by cytotoxic T cells (CTLs) and specific for
 CC human tumours, cancers and infected cells. The method involves screening
 CC the products of an expression library generated from DNA/RNA of a cell
 CC expressing a target epitope with cytotoxic T cells generated against
 CC the cell to identify DNA clones expressing target epitope or providing
 CC cytotoxic T cells specific for a gene product differentially expressed
 CC by a cell and measuring the cross-reactivity of the cytotoxic T cells
 CC for cells expressing a target epitope in which the target epitope is
 CC identified as a gene product inducing cytotoxic T cells. The method is
 CC useful for identifying a target epitope or antigen specific for a tumour
 CC cell. The target epitope is also useful for identifying target antigens
 CC in other target cells against which it is desirable to induce cell-
 CC mediated immunity. The antigen identified by the method is useful
 CC in immunogenic compositions and vaccine preparations to induce the
 CC regression of tumours, cancers and infections in mammals. The invention
 CC also relates to vaccinia viral vectors which are useful for treating
 CC tumour-bearing mammals, including humans to generate immune response
 CC against the tumour cells. They are also useful for immunising or
 CC vaccinating tumour-free subjects to prevent tumour formation. The
 CC present sequence is cytotoxic T-cell epitope for ovalbumin. This
 CC peptide is used in the exemplification of the invention.
 XX
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB 1 SIINFEKL 8
 RESULT 56
 ABB09907
 ID ABB09907 standard; peptide; 8 AA.
 XX
 AC ABB09907;
 XX

DT 10-JUN-2002 (first entry)
 XX
 DE Ovalbumin peptide (H-2kb CTL epitope).
 XX
 KW Ovalbumin; cytokine; ELISPOT assay; polycationic substance;
 KW cytokine secreting cell; interleukin; IL; interferon; IFN; TNF; CSF;
 KW tumour necrosis factor; colony stimulating factor;
 KW enzyme-linked immunosorbent spot assay.
 XX
 OS Aves.
 XX
 PN WO200179854-A1.
 XX
 PD 25-OCT-2001.
 XX
 PF 12-APR-2001; 2001WO-EP04208.
 XX
 PR 13-APR-2000; 2000AT-0000645.
 XX
 PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 XX
 PI Schalich J, Buschle M;
 XX
 DR WPI; 2002-240136/29.
 XX
 CC Identification of cytokine secreting cells involves incubating cell
 PT suspension containing antigen specific cells in the presence of
 PT polycationic substance -
 XX
 PS Example 1; Page 10; 22pp; English.
 XX
 CC The sequence represents an ovalbumin peptide (H-2kb CTL epitope from
 CC ovalbumin). The invention relates to a novel method for identifying
 CC cytokine secreting cells (especially ELISPOT assays) using a polycationic
 CC substance. The method is useful for identifying cytokine secreting cells.
 CC The cytokine may be selected from interleukins (IL), interferons (IFN),
 CC tumour necrosis factors (TNF), and colony stimulating factors (CSF).
 XX
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB 1 SIINFEKL 8
 RESULT 57
 AAU76942
 ID AAU76942 standard; Peptide; 8 AA.
 XX
 AC AAU76942;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE OVA peptide (257-264).
 XX
 KW ODN; immunostimulatory oligodeoxynucleotide; Ovalbumin; OVA;
 KW cardiant; vaccine; tuberculosis; diphtheria; pertussis; measles;
 KW tetanus; acquired immune deficiency syndrome; AIDS; malaria;
 KW cardiovascular disease; cancer; deoxyinosine; chicken.
 XX
 OS Gallus sp.
 XX
 PN WO200193905-A1.
 XX
 PD 13-DEC-2001.
 XX
 PF 07-JUN-2001; 2001WO-EP06433.
 XX
 PR 08-JUN-2000; 2000AT-0001000.

```

PR 23-NOV-2000; 2000AT-0001973.
XX
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX Schmidt W, Lingnau K, Schellack C, Egyed A;
XX
XX WPI; 2002-240927/29.
XX
XX New oligodeoxynucleic acid molecule useful as an immunostimulatory
XX agent
XX
XX Example 1; Page 18; 52pp; English.
XX
XX This invention relates to immunostimulatory oligodeoxynucleic acid
XX molecules (ODN) that can be used to enhance an immune response for
XX use in vaccines. The immunostimulatory oligonucleotides of the invention
XX and pharmaceutical compounds containing them may be used as medicine,
XX especially as an immunostimulatory agent, for the preparation of
XX vaccines useful for the treatment of tuberculosis, diphtheria,
XX pertussis, measles and tetanus, acquired immune deficiency syndrome
XX (AIDS), malaria, cardiovascular diseases, and cancer. Oligonucleotides
XX containing deoxyinosine residues (I-ODN) show a better immunostimulatory
XX effect compared to prior art compounds containing CpG motifs. The ODNs
XX of the invention produce more specific immune response to a given
XX antigen or antigen fragment than the prior art compounds containing CpG.
XX Using immunostimulatory oligonucleotides containing deoxyinosine reduces
XX the induction of adverse side reactions, especially the induction of
XX systemic TNF-alpha or interleukin-6. The immunostimulatory effect of the
XX composition containing a polycationic polymer and an antigenic fragment
XX was significantly higher than could be expected from the addition of the
XX effects of each single component or even the addition of the effects of
XX the ODN or the polycation with the antigen. The present sequence
XX represents the ovalbumin (OVA) peptide used as an antigen in
XX examples of the method of the invention.
XX
XX Sequence 8 AA;
SQ
    Query Match          100.0%; Score 38; DB 23; Length 8;
    Best Local Similarity 100.0%; Pred. NO. 9.3e+05;
    Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    Qy 1 SIINFPEKL 8
    Db 1 SIINFPEKL 8
    RESULT 58
    AAU76802
    ID AAU76802 standard; Peptide; 8 AA.
    AC
    AC AAU76802;
    XX
    XX 21-MAY-2002 (first entry)
    DE
    DE MHC class I-restricted epitope of chicken ovalbumin.
    XX
    XX Chicken; ovalbumin; immunostimulant; T cell epitope; inosine; cytosine;
    KW polycationic peptide; systemic immune response; MHC class I; vaccine;
    KW major histocompatibility complex class I.
    XX
    XX Gallus gallus.
    OS
    OS WO200193903-A1.
    XX
    XX WO200193903-A1.
    XX
    XX 13-DEC-2001.
    XX
    XX 07-JUN-2001; 2001WO-EP06437.
    XX
    XX 08-JUN-2000; 2000AT-0001000.
    XX
    XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
    PA
    PA Egyed A, Lingnau K, Mattner F, Buschle M, Schmidt W;
    PI

```

```

XX
XX WPI; 2002-205813/26.
XX
XX Pharmaceutical composition for the preparation of vaccine comprises T cell
XX epitope(s) or its mixture, a polycationic peptide and a nucleic acid
XX based on inosine and cytosine.
XX
XX Example 1; Page 10; 45pp; English.
XX
XX The invention relates to a pharmaceutical composition comprising a T cell
XX epitope(s) or its mixture, a polycationic peptide and a nucleic acid
XX based on inosine and cytosine. The composition of the invention induces a
XX systemic immune response and is used for the preparation of a vaccine.
XX This sequence represents an MHC class I-restricted epitope of chicken
XX ovalbumin, used to test enhancement of immune response against an
XX ovalbumin-derived peptide.
XX
XX Sequence 8 AA;
SQ
    Query Match          100.0%; Score 38; DB 23; Length 8;
    Best Local Similarity 100.0%; Pred. NO. 9.3e+05;
    Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    Qy 1 SIINFPEKL 8
    Db 1 SIINFPEKL 8
    RESULT 59
    AAU76869
    ID AAU76869 standard; Peptide; 8 AA.
    AC
    AC AAU76869;
    XX
    XX 21-MAY-2002 (first entry)
    DT
    DT OVA peptide fragment.
    DE
    DE Membrane vesicle; exosome; density cushion centrifugation;
    KW dendritic cell; MHC; major histocompatibility complex; CDI; tumour;
    KW immunotherapy treatment; cancer; infection; immune disease; antitumour;
    KW cytostatic; OVA.
    XX
    XX Unidentified.
    OS
    OS WO200182958-A2.
    XX
    XX 08-NOV-2001.
    PD
    PD 11-APR-2001; 2001WO-EP04173.
    XX
    XX 27-APR-2000; 2000US-0561205.
    PR
    PR 09-FEB-2001; 2001US-0780748.
    XX
    XX (APCE-) AP CELLS INC.
    PA
    PA Lamparski H, Ruegg C, Le Pecq J, Hsu D, Yao J;
    XX
    XX WPI; 2002-066489/09.
    DR
    DR Preparing membrane vesicle from biological sample for treating cancer,
    PT by culturing membrane vesicle-producing cells to release vesicles,
    PT enriching vesicles and subjecting sample to density cushion
    PT centrifugation
    XX
    XX Example 12; Page 55; 103pp; English.
    PS
    PS The invention relates to a method for preparing membrane vesicles (in
    CC particular exosomes) from a biological sample, comprising culturing a
    CC population of membrane vesicle-producing cells under conditions allowing
    CC the release of the vesicles, enriching the vesicles and treating the
    CC enriched biological sample by density cushion centrifugation. Immunogenic
    CC membrane vesicles are useful for producing an immune response in a

```

CC subject, by obtaining a biological sample containing dendritic cells,
 CC isolating or purifying a membrane vesicle from the sample, contacting the
 CC purified vesicle with a peptide or a lipid under conditions allowing the
 CC peptide or lipid to bind an MHC or CD1 molecule at the surface of the
 CC vesicle, and administering the vesicle to the subject. Membrane vesicles
 CC are useful for immunotherapy treatment or prophylaxis of tumours, and for
 CC treating various disease conditions such as cancer, infections, and
 CC immune diseases. This sequence represents an OVA peptide fragment which
 CC is contacted with an exosome, to create an immunogenic membrane vesicle.

XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINPEKL 8
 |||||
 Db 1 SIINPEKL 8

RESULT 60
 AAU11866
 ID AAU11866 standard; Peptide; 8 AA.

XX AC AAU11866;

XX DT 26-MAR-2002 (first entry)

XX DE Ovalbumin driven control peptide.

XX KW T0 terminator; pGA; DNA vaccine; anti-HIV; virucide;
 KW Human Immunodeficiency Virus; HIV; Gag; HIV gp120; HIV Pol; HIV Env;
 KW HIV VLP; measles fusion protein; measles haemagglutinin; epitope;
 KW measles nucleoprotein; influenza haemagglutinin; C3d gene; ovalbumin;
 KW cell-mediated immune response; humoral immune response; infection.

XX OS Unidentified.

XX PN WO200192470-A2.

XX PD 06-DEC-2001.

XX PF 02-MAR-2001; 2001WO-US06795.

XX PR 02-MAR-2000; 2000US-186364P.

XX PR 01-DEC-2000; 2000US-251083P.

XX PA (UYEM-) UNIV EMORY.

XX PI Robinson HL, Smith JM, Ross TM, Bright RA, Hua J, Ellenberger D;

XX DR WPI; 2002-075465/10.

XX PT Novel pGA vector useful for immunising patient against measles,
 PT Influenza has termination sequence encoding lambda T0 terminator and a
 PT eukaryotic transcription cassette with vaccine insert encoding
 PT immunogens of pathogens -

XX PS Example 14; Page 61; 174pp; English.

XX CC The invention relates to a vector (a pGA construct) comprising a
 CC termination sequence coding for the lambda T0 terminator, a prokaryotic
 CC origin of replication, a selectable marker gene and a eukaryotic
 CC transcription cassette comprising a vaccine insert encoding one or more
 CC immunogens derived from a pathogen e.g. Human Immunodeficiency Virus
 CC (HIV) Gag, HIV gp120, HIV Pol, HIV Env, HIV VLP, or its mutants, measles
 CC fusion protein, measles haemagglutinin, measles nucleoprotein, influenza
 CC haemagglutinin, or its mutants, or subsequences, and optionally at least
 CC one C3d gene, is useful for immunising or treating a patient, when
 CC administered by an intramuscular or intradermal route. The immunisation
 CC methods using pGA elicit both cell-mediated and humoral immune responses
 CC that may limit the infection, spread or growth of the pathogen and result

CC in protection against subsequent challenge against the pathogen. The
 CC terminator sequence present prevents read-through from the kanamycin
 CC cassette into vaccine sequences while the plasmid is being produced in
 CC bacteria. Prevention of transcriptional read-through stabilises vaccine
 CC insert sequences by limiting the exposure of secondary structures that
 CC can be recognised by bacterial endonucleases. The present sequence
 CC is an ovalbumin control peptide used in an experiment to measure the
 CC T-cell response in monkeys inoculated with a pGA vector carrying vaccinia
 CC virus genes.

XX SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINPEKL 8
 |||||
 Db 1 SIINPEKL 8

RESULT 61
 AAU11239
 ID AAU11239 standard; peptide; 8 AA.

XX AC AAU11239;

XX DT 12-MAR-2002 (first entry)

XX DE Immunodominant Kb-restricted Cytotoxic T lymphocyte epitope #2.

XX KW Cytostatic; vaccine; tetanus toxin; PrC; tumour; CTL;
 KW Cytotoxic T-lymphocyte; immunodominant Kb-restricted CTL epitope.

XX OS Unidentified.

XX PN WO200179510-A1.

XX PD 25-OCT-2001.

XX PF 17-APR-2001; 2001WO-GB01719.

XX PR 17-APR-2000; 2000GB-0009470.

XX PA (CANC-) CANCER RES VENTURES LTD.

XX PI Rice J, Stevenson F;

XX DR WPI; 2002-066370/09.

XX PT Nucleic acid construct, useful to immunise against various diseases
 PT including cancer, expresses the first domain of tetanus toxin PrC fused
 PT to a disease peptide antigen to provide a vaccine -

XX PS Disclosure; Page 25; 71pp; English.

XX CC The invention relates to a nucleic acid construct for delivery into
 CC living cells in vivo, to induce an immune response to a disease peptide
 CC antigen, where the construct directs expression of a fusion protein
 CC comprising the peptide antigen and the first domain of PrC. Also
 CC included are a nucleic acid vector comprising the above construct,
 CC a host cell comprising the above construct or vector and a method of
 CC producing a nucleic acid construct for inducing an immune response.
 CC The method comprises identifying a nucleic acid sequence encoding a
 CC disease peptide antigen comprising epitopes characteristic of the
 CC disease, cloning the nucleic acid sequence, introducing the cloned
 CC nucleic acid into a vector which allows the antigen to be expressed as a
 CC fusion with a first domain PrC from tetanus toxin, and optionally
 CC isolating the construct from the vector. The construct or vector is used
 CC as a vaccine to induce an immune response, particularly to tumour
 CC antigens. The present sequence is an immunodominant Kb-restricted
 CC cytotoxic T-lymphocyte (CTL) epitope suitable for inclusion in the
 CC vaccine of the invention.

XX SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||

Db 1 SIINFEKL 8

RESULT 62
 AAU09820
 ID AAU09820 standard; peptide; 8 AA.
 XX AC AAU09820;
 XX DT 14-FEB-2002 (first entry)
 XX DE Ovalbumin-derived class I H-2Kb restricted peptide (245).
 XX KW Ovalbumin-derived class I H-2Kb restricted peptide (245); vaccine;
 KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
 KW popliteal lymph node; spleen; immune response; systemic response.
 XX OS Unidentified.
 XX PN WO200178767-A2.
 XX PD 25-OCT-2001.
 XX PF 17-APR-2001; 2001WO-EP04313.
 XX PR 14-APR-2000; 2000AT-0000657.
 XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 XX PI Mattner F, Zauner W, Schmidt W, Buschle M;
 XX WPI; 2002-025970/03.
 XX PT Pharmaceutical preparation for use as a potent vaccine for inducing an
 PT improved immune response in a mammal, comprises a modified peptide -
 XX Example 1; Page 8; 18pp; English.

XX CC The invention relates to a pharmaceutical preparation comprising a
 CC modified peptide, which induces an improved immune response in a mammal
 CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
 CC negatively charged (Glu), one positively charged (Lys) amino acid) was
 CC rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
 CC Asp, respectively. Results showed that the addition of 4 negatively-
 CC charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
 CC this peptide (in combination with poly-L-arginine) able to induce a high
 CC amount of specific interferon (IFN)-gamma-producing T cells in the
 CC draining (popliteal) lymph node (local response) and in the spleen
 CC (systemic response). Thus, the addition of hydrophobic amino acids as
 CC well as the addition of negatively charged amino acids transforms the
 CC peptide SIINFEKL to a good inducer of specific T cells. The modified
 CC peptides of the pharmaceutical composition induce a stronger immune
 CC response in a mammal compared to wild type antigens. The present
 CC sequence represents ovalbumin-derived class I H-2Kb restricted peptide
 CC (245) used to produce the modified hydrophobic peptides described in
 CC the method of the invention.

XX SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8

Db 1 SIINFEKL 8
 |||||

RESULT 63
 AAEL13215
 ID AAEL13215 standard; peptide; 8 AA.
 XX AC AAEL13215;
 XX DT 12-FEB-2002 (first entry)
 XX DE Hen egg ovalbumin (OVA) peptide.
 XX KW Cytotoxic T lymphocyte; CTL; T cell; tumour load; cancer radiotherapy;
 KW immunostimulatory sequence oligonucleotide; ISS-ODN; chemotherapy;
 KW immunosuppression; transplantation; autoimmune disease; infection;
 KW acquired immune deficiency syndrome; AIDS; intracellular pathogen;
 KW cytomegalovirus; mycobacterial infection; Epstein-Barr virus;
 KW varicella zoster virus; human immunodeficiency virus; HIV; hen;
 KW ovalbumin; OVA.
 XX OS Unidentified.
 XX PN WO200172123-A1.
 XX PD 04-OCT-2001.
 XX PF 28-MAR-2001; 2001WO-US10118.
 XX PR 28-MAR-2000; 2000US-192537P.
 XX PR 11-MAY-2000; 2000US-203567P.
 XX PR 05-JUL-2000; 2000US-215895P.
 XX PA (REGC) UNIV CALIFORNIA.
 XX PA (VETE-) DEPT VETERANS AFFAIRS.
 XX PI Raz E, Cho HJ, Richman DD, Horner AA;
 XX WPI; 2002-010699/01.
 XX PT Increasing antigen-specific cytotoxic T lymphocyte activity in a CD4+ T
 PT cell deficient individual, useful to treat immunodeficiency and block
 PT HIV infection, comprises administering immunostimulatory nucleic acid
 PT -
 XX Example 1; Page 44; 91pp; English.

XX CC The present invention relates to a method for increasing antigen-specific
 CC cytotoxic T lymphocyte (CTL) activity in a CD4+ T cell-deficient
 CC individual, comprising administering an immunostimulatory sequence
 CC oligonucleotide (ISS-ODN). The immunostimulatory nucleic acids of the
 CC invention are used in CD4+ T cell-deficient individuals to decrease
 CC tumour load, to treat a primary or acquired immunodeficiency,
 CC particularly where the acquired immunodeficiency is temporary and due
 CC to cancer radiotherapy or chemotherapy or immunosuppression following
 CC bone marrow or organ transplantation, or autoimmune disease treatment,
 CC or is acquired immunodeficiency syndrome (AIDS). The nucleic acids may
 CC be used to treat a person at risk of becoming CD4+ T cell-deficient,
 CC particularly where someone at risk of cancer recurrence. They are also
 CC used to treat infection, particularly by an intracellular pathogen,
 CC especially one caused by cytomegalovirus, Mycobacterium tuberculosis,
 CC M. avium, Epstein-Barr virus, a fungus yeast, varicella zoster virus or
 CC human immunodeficiency virus (HIV). The present sequence is a hen egg
 CC ovalbumin (OVA) peptide, used in the exemplification of the invention.

XX SQ Sequence 8 AA;

Query Match 100.0%; Score 38; DB 23; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8

```

Db      |||||
      1 SIINFELK 8

RESULT 64
AAE13436
ID  AAE13436 standard; peptide; 8 AA.
XX
AC  AAE13436;
XX
DT  12-FEB-2002 (first entry)
XX
DE  Chicken ovalbumin major histocompatibility complex class I epitope.
XX
KW  Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW  major histocompatibility complex; MHC; therapy; immune response;
KW  malignancy; chicken.
XX
OS  Gallus gallus.
XX
PN  WO200179259-A1.
XX
PD  25-OCT-2001.
XX
PF  17-APR-2001; 2001WO-US12567.
XX
PR  17-APR-2000; 2000US-197462P.
XX
PA  (ROTH/) ROTHMAN J E.
PA  (MAYH/) MAYHEW M.
PA  (HOEM/) HOE M.
XX
PI  Rothman JE, Mayhew M, Hoe M;
XX
DR  WPI; 2002-017594/02.
XX
PT  A new antigenic complex comprising epitopes non-covalently joined to a
PT  heat shock protein by a molecular tether designated a javelin are
PT  useful to treat or prevent infectious disease or malignancy -
XX
PS  Example; Page 15; 47pp; English.
XX
CC  The present invention relates to an antigenic complex, comprising a
CC  number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC  a tethering molecule referred to as javelin which has affinity for the
CC  HSP under physiological conditions, where the epitopes are covalently
CC  joined to the tethering molecule and one epitope is major
CC  histocompatibility complex class I (MHC) and the other MHC class II. The
CC  antigenic complex is used to induce immune responses directed towards the
CC  treatment or prevention of infectious diseases and malignancies. The
CC  present sequence is chicken ovalbumin MHC class I epitope.
XX
SQ  Sequence 8 AA;

Query Match      100.0%; Score 38; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIINFELK 8
      |||||
Db      1 SIINFELK 8

RESULT 65
AAU75056
ID  AAU75056 standard; peptide; 8 AA.
XX
AC  AAU75056;
XX
DT  23-APR-2002 (first entry)
XX
DE  Ovalbumin antigenic peptide.
XX

```

```

KW  Ovalbumin; Th1; Th2; antigen; immunosuppressive; cytostatic;
KW  adjuvant; lipopolysaccharide; lipid; vaccine; immunogenicity;
KW  immunocompetence; autoimmune disease; infectious disease; OVA;
KW  graft-versus- host disease; tumour; transgenic T cell; chicken.
XX
OS  Gallus gallus.
XX
PN  WO200197838-A2.
XX
PD  27-DEC-2001.
XX
PF  18-JUN-2001; 2001WO-US19411.
XX
PR  16-JUN-2000; 2000US-212182P.
XX
PA  (BAYU ) BAYLOR RES INST.
XX
PI  Pulendran B, Banchereau JF, Cutler CW;
XX
DR  WPI; 2002-114543/15.
XX
PT  Use of adjuvants comprising isolated lipid groups such as Porphyromonas
PT  gingivalis lipopolysaccharides or its detoxified forms or derivatives
PT  for preparation of compositions to elicit T-helper cell responses
PT  mammals -
XX
PS  Disclosure; Page 20; 58pp; English.
XX
CC  This invention relates to the use of adjuvants comprising isolated lipid
CC  groups such as Porphyromonas gingivalis lipopolysaccharide, its
CC  detoxified forms or derivatives, for preparing compositions for
CC  eliciting Th2 responses, enhancing vaccine immunogenicity, modulating
CC  immunocompetence, treating autoimmune/infectious disease, stimulating
CC  interleukin-5 (IL-5)/IL-13 production or dampening interferon gamma
CC  production in mammal. The adjuvant or a pharmaceutical compound
CC  containing it is useful for enhancing antibody harvest in a laboratory
CC  animal through an elicited Th2 immune response, or by modulating Th2
CC  immune responses. The adjuvant is also useful to study the Th2
CC  immune response in laboratory animal research, for the treatment or
CC  prophylactic vaccination of humans or animals against graft-versus-
CC  host disease, and for treating tumours. The present sequence
CC  represents the Ovalbumin (OVA) peptide used as an antigen to
CC  stimulate transgenic T cells in the method of the invention.
XX
SQ  Sequence 8 AA;

Query Match      100.0%; Score 38; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIINFELK 8
      |||||
Db      1 SIINFELK 8

RESULT 66
ABU07743
ID  ABU07743 standard; Peptide; 8 AA.
XX
AC  ABU07743;
XX
DT  23-MAY-2003 (first entry)
XX
DE  Chicken ovalbumin epitope presented on murine Kb MHC class I.
XX
KW  Chicken; ovalbumin; OVA; cytolysin; vaccine delivery; Kb MHC class I;
KW  intracellular delivery vehicle; nonvirulent bacterium; drug delivery;
KW  gene therapy; biosynthesis; high level protein delivery;
KW  major histocompatibility complex; cytosolic protein delivery.
XX
OS  Gallus gallus.
XX
PN  US2002142007-A1.

```

XX PD 03-OCT-2002.
 XX PF 07-SEP-2001; 2001US-0949109.
 XX PR 13-AUG-1998; 98US-01333914.
 XX PR 21-DEC-1999; 99US-0469197.
 XX PA (PORT/) PORTNOY D A.
 XX PA (HIGG/) HIGGINS D E.
 XX PI Portnoy DA, Higgins DE;
 XX PI WPI; 2003-328328/31.
 XX PT New nonvirulent bacterium with genes coding for a non-secreted foreign
 PT cytolyisin or a different foreign agent, useful as an intracellular
 PT delivery vehicle for delivering, e.g. vaccines, drugs or genes for
 PT therapy to eukaryotic cells -
 XX PS Example; Page 6; 14pp; English.
 XX CC The invention relates to a nonvirulent bacterium, which comprises a first
 CC gene encoding a non-secreted foreign cytolyisin operably linked to a
 CC heterologous promoter and a second gene encoding a different foreign
 CC agent. The nonvirulent bacterium is useful as an intracellular delivery
 CC vehicle, particularly of agents to eukaryotic cells. The nonvirulent
 CC bacterium is particularly useful for delivering foreign agents for
 CC diagnosis, therapy (e.g. prophylactics such as vaccine, delivery of
 CC therapeutic drug, or gene therapy) or biosynthesis. The nonvirulent
 CC bacterium is also useful for delivering nucleic acids that provide
 CC templates for transcription or translation, or provide modulators of
 CC transcription and/or translation. No protein purification is required
 CC compared to prior art delivery systems. In addition, high levels of
 CC protein can be delivered to the cytosol of virtually any cell and the
 CC levels can be controlled through the use of inducible promoters. L.
 CC monocytogenes LLO (listeriolysin or cytolyisin) was transformed in
 CC E.coli cells and used as a system to deliver chicken ovalbumin to the
 CC cytosol of macrophages. The present sequence represents the chicken
 CC ovalbumin epitope presented on murine Kb major histocompatibility
 CC complex, MHC, class I.
 XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 24; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 Db 1 SIINFEKL 8
 RESULT 67
 ABU08619
 ID ABU08619 standard; Peptide; 8 AA.
 XX AC ABU08619;
 XX DT 23-MAY-2003 (first entry)
 XX DE Ovalbumin (OVA) residues 257-264.
 XX KW Cancer; tumour; antigen-presenting cell; APC; tumour cell conjugate;
 KW cytokine; interleukin; interferon; IFN alpha; IFNbeta; IFNgamma;
 KW tumour necrosis factor; TNF; transforming growth factor; TGF;
 KW granulocyte-macrophage colony stimulating factor; GM-CSF; vaccine;
 KW melanoma; kidney cancer; pulmonary carcinoma; hepatic carcinoma;
 KW mammary cancer; prostatic carcinoma; gastric carcinoma; leukaemia;
 KW ovalbumin-specific tetramer; (OVA)-specific tetramer; ovalbumin;
 XX OVA.
 XX OS Gallus gallus.

XX PN US2002182194-A1.
 XX PD 05-DEC-2002.
 XX PF 25-MAR-2002; 2002US-0106173.
 XX PR 04-APR-2001; 2001CN-0105852.
 XX PA (SHAN-) SHANGHAI BRILLIANCE BIOTECH INST.
 XX PI Ju D, Tao Q, Ye D;
 XX PI WPI; 2003-328591/31.
 XX DR New antigen-presenting cell and tumor cell conjugates, where the
 PT antigen-presenting cell is modified by a cytokine gene, useful for the
 PT preparation of a medicine for the therapy of cancer or a vaccine for
 PT the prophylaxis of cancer -
 XX PS Example 1; Page 4; 23pp; English.
 XX CC The invention describes an antigen-presenting cell (APC)/tumour cell
 CC conjugate, where the APC is modified by a cytokine gene (interleukin
 CC (IL-2, IL-3, IL-4, IL-6, IL-12, IL-18, interferon (IFN) alpha, IFNbeta,
 CC IFNgamma, tumour necrosis factor (TNF), transforming growth factor (TGF)
 CC and/or granulocyte-macrophage colony stimulating factor (GM-CSF)). The
 CC antigen-presenting cell/tumour cell conjugate is useful for the
 CC preparation of a medicine for the therapy of cancer or a vaccine for the
 CC prophylaxis of cancer. The cancer includes melanoma, kidney cancer,
 CC pulmonary carcinoma, hepatic carcinoma, mammary cancer, prostatic
 CC carcinoma, gastric carcinoma and leukaemia. This is the amino acid
 CC sequence of Ovalbumin (OVA) residues 257-264 used in the creation of
 CC Ovalbumin (OVA)-specific tetramers used in the vaccine of the invention.
 XX SQ Sequence 8 AA;
 Query Match 100.0%; Score 38; DB 24; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 Db 1 SIINFEKL 8
 RESULT 68
 ABP57401
 ID ABP57401 standard; peptide; 8 AA.
 XX AC ABP57401;
 XX DT 23-APR-2003 (first entry)
 XX DE Synthetic 8mer peptide.
 XX KW Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
 KW Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
 KW virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
 XX OS Synthetic.
 XX PN WO2003000899-A1.
 XX PD 03-JAN-2003.
 XX PF 20-JUN-2002; 2002WO-GB02829.
 XX PR 22-JUN-2001; 2001GB-0015382.
 XX PA (UYBR-) UNIV BRISTOL.
 XX PI Hirst TR;

```

XX WPI; 2003-175291/17.
XX
XX Use of a mutant form of B subunit of Escherichia coli heat labile
XX enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX target cell for treating viral infection or cancer
XX
XX Example 5; Page 45; 84pp; English.
XX
XX The present invention describes a mutant form of B subunit of Escherichia
XX coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX from Vibrio cholerae which is useful for delivering an agent to a target
XX cell, and has GM-1 ganglioside receptor binding activity but has reduced
XX immunogenic and immunomodulatory activity relative to the wild-type form
XX of EtxB or CtxB. Also described: (1) treating a disease or condition in
XX a subject; (2) delivering the agent using the mutant to a target cell;
XX (3) a composition; and (4) a kit for delivering the agent to a target
XX cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
XX can be used in vaccines. The mutant can be used for the preparation of
XX a medicament for delivering an exogenous peptide, which is the agent,
XX into the major histocompatibility complex (MHC) Class I antigen
XX processing and presenting pathways to elicit a cytotoxic T lymphocyte
XX (CTL) response, or for separate, simultaneous or combined use for
XX treating viral infection or cancer. The mutant form of EtxB or CtxB
XX enters mammalian cells without inducing a potent anti-B-subunit response
XX and immunomodulatory response. It may be linked with an agent to
XX upregulate the presentation of the antigen or antigenic determinant.
XX The present sequence represents a peptide which is used in an example
XX from the present invention.
XX
XX Sequence 8 AA;
XX
XX Query Match 100.0%; Score 38; DB 24; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SIINFELK 8
XX Db |||||
XX 1 SIINFELK 8
XX
XX RESULT 69
XX ABP58359
XX ID ABP58359 standard; Peptide; 8 AA.
XX AC ABP58359;
XX DT 07-APR-2003 (first entry)
XX DE Ovalbumin-derived peptide OVA257-264.
XX KW Chicken; ovalbumin; allergen; immunostimulant;
XX KW oligodeoxynucleic acid; ODN; vaccine.
XX OS Gallus sp.
XX FN WO200295027-A2.
XX PD 28-NOV-2002.
XX PF 17-MAY-2002; 2002WO-EP05448.
XX PR 21-MAY-2001; 2001AT-0000805.
XX PA (INTE-) INTERCELL BIOMEDIZINISCHE FORSCHUNGS.
XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX PI Lingnau K, Schellack C, Schmidt W;
XX WPI; 2003-183880/18.
XX
XX New oligodeoxynucleic acid molecules useful for the preparation of
XX vaccine -

```

```

XX Example 3; Page 23; 57pp; English.
XX
XX The present sequence is that of OVA257-264, a major
XX histocompatibility complex class I (H-2Kb)-restricted epitope of
XX chicken ovalbumin. An example from the invention describes the
XX generation of specific immune responses against this allergen-derived
XX peptide using deoxyuridine monophosphate-modified oligonucleotide
XX U-ODN 13 (see AB224776). U-ODN 13 is an example of new
XX oligodeoxynucleic acid (ODN) molecules useful in the preparation of
XX vaccines. The invention is based on the discovery that ODNs
XX containing deoxyuridine residues have an immunostimulatory effect
XX comparable to, or greater than, ODNs containing CpG motifs.
XX combining the ODN with an antigen strongly increases the potential
XX of the antigen to raise the protection/immune response of a
XX vaccinated individual.
XX
XX Sequence 8 AA;
XX
XX Query Match 100.0%; Score 38; DB 24; Length 8;
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SIINFELK 8
XX Db |||||
XX 1 SIINFELK 8
XX
XX RESULT 70
XX ABG73081
XX ID ABG73081 standard; Peptide; 8 AA.
XX AC ABG73081;
XX DT 02-APR-2003 (first entry)
XX DE MHC Class I peptide OVA-8.
XX KW Antigen-specific T lymphocyte; MHC-antigen complex; MHC Class I peptide;
XX KW major histocompatibility complex; tumour-specific killer T cell;
XX KW virus-specific killer T cell; cytostatic; virucide; OVA-8.
XX OS Synthetic.
XX PN US2002151690-A1.
XX PD 17-OCT-2002.
XX PF 05-NOV-1999; 99US-0434965.
XX PR 12-AUG-1997; 97US-0909549.
XX PA (LUXE/) LUXEMBURG A T.
XX PA (JACK/) JACKSON M R.
XX PA (PETE/) PETER P A.
XX PI Luxembourg AT, Jackson MR, Peter PA;
XX PD WPI; 2003-182532/18.
XX
XX Enriching antigen-specific T lymphocytes, for purifying or expanding in
XX vitro tumour- or virus-specific killer T cells for cell therapy,
XX comprises capture of the lymphocytes on a substrate coated with
XX antigenic peptide-MHC complexes
XX
XX Example 2; Page 5; 40pp; English.
XX
XX The invention relates to a method for enriching antigen-specific T
XX lymphocytes, comprising contacting a heterogeneous population of
XX antigen-specific T lymphocytes with a matrix comprising MHC-antigen
XX complexes for a period of time sufficient to allow the antigen-specific T
XX lymphocytes to interact with the matrix, and eluting the antigen-specific
XX T lymphocytes from the matrix to provide an enriched population of

```


CC antigen-specific T lymphocytes. The MHC-antigen complexes comprise one or
 CC more antigens. Also claimed is a matrix for capturing antigen-specific T
 CC lymphocytes, comprising a support having on its surface an immobilised
 CC Class I peptide and a predetermined amount of an antigen, or for
 CC capturing antigens, comprising a support having on its surface an
 CC immobilised empty Class I peptide which is capable of binding one or more
 CC antigens, and isolating antigen-specific T lymphocytes from a
 CC heterogenous population of cells from a patient. The methods are useful
 CC for enriching antigen-specific T lymphocytes to purify and expand in
 CC vitro tumour- and virus-specific killer T cells for cell therapy. The
 CC methods are also useful for isolating or preparing a population of the
 CC antigen-specific T lymphocytes from a patient for treatment of the
 CC patient's disease or condition. This sequence represents an MHC Class I
 CC peptide used in the method of the invention.

XX Sequence 8 AA;
 SQ
 Query Match 100.0%; Score 38; DB 24; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFELK 8
 |||||
 Db 1 SIINFELK 8

RESULT 71
 ABP56760
 ID ABP56760 standard; peptide; 8 AA.

AC ABP56760;
 DT 31-MAR-2003 (first entry)

XX Ovalbumin derived peptide OVA 257-264.

DE Stabilisation; polycationic polymer; medicine; vaccination;
 KW gene therapy; drug; ovalbumin.

XX Gallus sp.
 OS Synthetic.

XX WO200294845-A2.

XX 28-NOV-2002.

XX 17-MAY-2002; 2002WO-EP05447.

XX 21-MAY-2001; 2001AT-0000805.

XX (INTE-) INTERCELL BIOMEDIZINISCHE FORSCHUNGS.
 XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.

XX Schellack C, Lingnau K, Schmidt W;

XX WPI; 2003-140356/13.

XX Use of polycationic polymer for stabilization of nucleic acids -

XX Example 1; Page 6; 28pp; English.

XX The present invention describes a method for the stabilisation of nucleic
 CC acids involving contacting nucleic acids with a polycationic polymer in
 CC aqueous solution or suspension. The method can be used for the
 CC stabilisation of nucleic acid which can be used in medicines e.g.
 CC vaccinations, as a general immunostimulant, as an antitense drug
 CC or gene therapy drug. The present sequence represents an ovalbumin
 CC derived peptide, designated OVA 257-264, which is used in an example
 CC from the present invention.

XX Sequence 8 AA;
 SQ
 Query Match 100.0%; Score 38; DB 24; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFELK 8
 |||||
 Db 1 SIINFELK 8

RESULT 72
 ABP60027
 ID ABP60027 standard; Peptide; 8 AA.

XX AC ABP60027;

XX DT 07-MAR-2003 (first entry)

XX DE Ovalbumin antigenic peptide.

XX KW TOP; thimet oligopeptidase; EC3.4.25.15; cytostatic; tumour;
 KW immunostimulant; major histocompatibility complex class I; MHC;
 KW T-cell immunity; chicken.

XX OS Gallus gallus.

XX FN WO200279388-A2.

XX PD 10-OCT-2002.

XX PF 01-APR-2002; 2002WO-US10385.

XX PR 30-MAR-2001; 2001US-280669P.

XX PA (UYMA-) UNIV MASSACHUSETTS.

XX PI Rock KL, Goldberg AL;

XX DR WPI; 2003-103265/09.

XX PT New recombinant cell comprising an exogenously derived nucleic acid
 PT coding for a thimet oligopeptidase polypeptide, useful for modulating
 PT an antigenic response in a mammal for treating e.g., tumour -

XX Example 1; Page 50; 73pp; English.

XX The invention relates to a new recombinant cell comprising an exogenously
 CC derived nucleic acid that codes for a thimet oligopeptidase (TOP)
 CC polypeptide. The TOP polypeptide is overexpressed in the cell compared to
 CC a wild-type cell from which the recombinant cell is derived. The activity
 CC of TOP may be described as cytostatic and immunostimulatory. Thimet
 CC oligopeptidase (TOP, EC3.4.25.15) plays a key role in modulating levels
 CC of major histocompatibility complex (MHC) class I-presented peptides. The
 CC recombinant host cell of the invention is useful for modulating an
 CC antigenic response in a mammal. Methods of the invention are useful for
 CC screening a test compound for its ability to serve as an immunomodulatory
 CC agent and identifying an antigen resistant to thimet oligopeptidase
 CC degradation. A method of the invention is useful for increasing CD8
 CC T-cell immunity, which uses vaccination with a TOP inhibitor for
 CC decreasing TOP expression or activity. The vaccination method uses
 CC treated tumour cells, antigen bearing/pulsed dendritic cells or injection
 CC of a viral vector. The recombinant host cell is useful for treating
 CC tumours. The current sequence represents an ovalbumin antigenic peptide
 CC that is used in an example from the invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 38; DB 24; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFELK 8
 |||||
 Db 1 SIINFELK 8

RESULT 73
ABU11029
ID ABU11029 standard; Peptide; 8 AA.
XX AC ABU11029;
XX DT 04-FEB-2003 (first entry)
XX DE Ovalbumin immunodominant epitope.
XX KW Epitope; antigen-specific immunoglobulin; Ig; early/late promoter;
KW heavy chain constant region; light chain constant region;
KW variable region; camelised Ig heavy chain variable region; MHC; CTL;
KW major histocompatibility class; cytotoxic T-lymphocyte.
XX OS Unidentified.
XX PN US2002123057-A1.
XX PD 05-SEP-2002.
XX PF 14-NOV-2001; 2001US-0987456.
XX PR 17-NOV-2000; 2000US-249268P.
XX PR 18-JAN-2001; 2001US-262067P.
XX PR 27-FEB-2001; 2001US-271424P.
XX PR 15-JUN-2001; 2001US-298087P.
XX PA (UYRP) UNIV ROCHESTER.
XX PI Zauderer M, Smith ES;
XX DR WPI; 2003-066785/06.
XX PT Selecting polynucleotides which encode antigen-specific immunoglobulin
PT molecules, by introducing the library of polynucleotides into the host
PT cells, and recovering the polynucleotides of the library for the
PT antigen -
XX Example 5; Page 45; 108pp; English.
XX The invention relates to selecting polynucleotides which encode antigen
CC -specific immunoglobulins (Ig) (or fragments) comprising introducing into
CC a population of host cells, a 1st and 2nd library of polynucleotides
CC encoding, several 1st and 2nd Ig subunit polypeptides, permitting
CC expression of Ig molecules (via control element e.g. an early/late
CC promoter), contacting Ig molecules with an antigen,
CC and recovering polynucleotides of the 1st library for the antigen.
CC The Ig molecules are heavy and light chain constant regions and
CC variable regions linked via peptide linkers and optionally directed via
CC signal peptides or transmembrane domains to different cell compartments.
CC Also included is a method of selecting polynucleotides which encode a
CC single-domain antigen-specific Ig molecule (its anti-specific fragment),
CC by: (a) introducing into a population of eukaryotic host cells capable of
CC expressing the Ig molecule a library of polynucleotides encoding
CC (through operable association with a transcriptional control region)
CC several single-domain Ig polypeptides (each comprising a Ig heavy chain
CC constant region, a camelised Ig heavy chain variable region, and a
CC signal peptide capable of directing cell surface expression or
CC secretion of Ig subunit polypeptide); (b) permitting expression of Ig
CC molecules (or antigen-specific fragments) from the host cells;
CC (c) contacting the Ig molecules with an antigen; and (d) recovering
CC polynucleotides of the library from those host cells expressing Ig
CC molecules which bind the antigens. The methods are useful for selecting
CC polynucleotides which encode an antigen-specific Ig molecule, or its
CC fragment. The present sequence is a major histocompatibility class II
CC (MHC II), cytotoxic T-lymphocyte (CTL) epitope expressed on the surface
CC of host cells used in the method of the invention.
XX SQ Sequence 8 AA;
XX Query Match 100.0%; Score 38; DB 24; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 1 SIINFEKL 8
RESULT 74
AAB84323
ID AAB84323 standard; peptide; 9 AA.
XX AC AAB84323;
XX DT 22-AUG-2001 (first entry)
XX DE Peptide used to produce lemA peptide variants.
XX KW lemA; CD8+ epitope; T cell response.
XX OS Synthetic.
XX PN WO200140275-A2.
XX PD 07-JUN-2001.
XX PF 06-DEC-2000; 2000WO-US33027.
XX PR 06-DEC-1999; 99US-0169227.
XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX PI Kurlander RJ, Chao E, Fields J;
XX DR WPI; 2001-389952/41.
XX PT New isolated variant of lemA, tleMA, comprising a hydrophobic element
PT joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
PT response or as a treatment or prophylactic against diseases -
XX PS Disclosure; Page 7; 65pp; English.
XX The specification describes a peptide variant of lemA, comprising a
CC hydrophobic element joined to a CD8+ epitope. The peptides may be
CC used therapeutically by administering the peptides to a patient having
CC a need to induce a directed CD8+ T cell response. The peptide may also
CC be used as a preventive measure to avoid a disease or condition, or to
CC treat subjects already afflicted with a disease. The present sequence
CC was used to create peptides of the invention.
XX SQ Sequence 9 AA;
XX Query Match 100.0%; Score 38; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SIINFEKL 8
DB 2 SIINFEKL 9
RESULT 75
ABP57402
ID ABP57402 standard; peptide; 9 AA.
XX AC ABP57402;
XX DT 23-APR-2003 (first entry)
XX DE Synthetic 9mer peptide.
XX KW Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;

KW Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
 XX virucide; cytostatic; vaccine; viral infection; cancer; EtXB; CtxB.
 XX Synthetic.
 OS
 XX WO2003000899-A1.
 XX
 XX 03-JAN-2003.
 XX
 XX 20-JUN-2002; 2002WO-GB02829.
 XX
 XX 22-JUN-2001; 2001GB-0015382.
 XX (UYBR-) UNIV BRISTOL.
 XX
 XX Hirst TR;
 XX
 XX WPI; 2003-175291/17.
 XX
 XX Use of a mutant form of B subunit of Escherichia coli heat labile
 PT enterotoxin or B subunit of cholera toxin for delivering an agent to a
 PT target cell for treating viral infection or cancer -
 XX
 XX Example 5; Page 45; 84pp; English.
 PS
 XX The present invention describes a mutant form of B subunit of Escherichia
 CC coli heat labile enterotoxin (EtXB) or B subunit of cholera toxin (CtxB)
 CC from Vibrio cholerae which is useful for delivering an agent to a target
 CC cell, and has GM-1 ganglioside receptor binding activity but has reduced
 CC immunogenic and immunomodulatory activity relative to the wild-type form
 CC of EtXB or CtxB. Also described: (1) treating a disease or condition in
 CC a subject; (2) delivering the agent using the mutant to a target cell;
 CC (3) a composition; and (4) a kit for delivering the agent to a target
 CC cell. Mutant EtXB and CtxB have virucide and cytostatic activities and
 CC can be used in vaccines. The mutant can be used for the preparation of
 CC a medicament for delivering an exogenous peptide, which is the agent,
 CC into the major histocompatibility complex (MHC) Class I antigen
 CC processing and presenting pathways to elicit a cytotoxic T lymphocyte
 CC (CTL) response, or for separate, simultaneous or combined use for
 CC treating viral infection or cancer. The mutant form of EtXB or CtxB
 CC enters mammalian cells without inducing a potent anti-B-subunit response
 CC and immunomodulatory response. It may be linked with an agent to
 CC upregulate the presentation of the antigen or antigenic determinant.
 CC The present sequence represents a peptide which is used in an example
 CC from the present invention.
 XX
 XX Sequence 9 AA;
 SQ
 Query Match 100.0%; Score 38; DB 24; Length 9;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFELK 8
 Db |||||
 2 SIINFELK 9
 RESULT 76
 AAW04643
 ID AAW04643 standard; peptide; 10 AA.
 XX
 XX AAW04643;
 XX
 XX 01-AUG-1997 (first entry)
 XX
 XX Ovalbumin-derived activated CD8+ T cells epitope OVA10N.
 DE
 XX Macrophage; artificial antigen presenting cell; APC; cancer;
 KW tumours; neoplasia; viral infection; retroviral infection;
 KW autoimmune.
 XX
 XX Synthetic.
 OS

PN WO9637107-A1.
 XX
 PD 28-NOV-1996.
 XX
 XX 22-MAY-1996; 96WO-US07436.
 XX
 XX 23-MAY-1995; 95US-0447761.
 XX
 XX (SCRI) SCRIPPS RES INST.
 XX
 XX DeBruijn MLH, Jackson MR, Peterson PA;
 XX
 XX WPI; 1997-020850/02.
 XX
 XX Prodn. of activated CD8+ T cells directed to specific antigen - can
 PT specifically kill target cells useful to treat, e.g. cancer
 PT
 XX
 XX Example 1; Page 26; 84pp; English.
 PS
 XX The method for the production of activated CD8+ T cells specifically
 CC directed towards a particular antigen involves affixing peptides
 CC corresponding to the particular antigen to an artificial support;
 CC contacting macrophages with the affixed peptides for a time sufficient
 CC for the peptides to be engulfed, and at least a portion of the peptides
 CC to be presented on the surface of the macrophage; and contacting
 CC unprimed CD8+ T cells with the peptide presenting macrophages for a
 CC time sufficient to activate the unprimed CD8+ T cells. The present
 CC sequence represents a peptide designated OVALON which corresponds to
 CC ovalbumin, a Kb-restricted peptide antigen. This represents the optimal
 CC peptide with the addition of two amino acids at th eamino-terminus.
 CC Small extensions to the optimal peptide affect the affinity of the
 CC peptide for soluble class I molecules in vitro e.g. the addition
 CC of two amino acids to the amino-terminus lowers the affinity to Kb by
 CC 76-fold compared to the optimal peptide; addition of two amino acids to
 CC the carboxy-terminus lowers the affinity by 4-fold. The method,
 CC macrophages and artificial antigen presenting cell, having a peptide
 CC corresponding to an artificial antigen present on its surface and at
 CC least a portion of an artificial support in its interior, can be used to
 CC treat conditions (e.g. cancer, tumours, neoplasia, viral or retroviral
 CC infection or autoimmune or autoimmune-type conditions) in patients via
 CC the specific killing of target cells.
 XX
 XX Sequence 10 AA;
 SQ
 Query Match 100.0%; Score 38; DB 18; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFELK 8
 Db |||||
 3 SIINFELK 10
 RESULT 77
 AAW04644
 ID AAW04644 standard; peptide; 10 AA.
 XX
 XX AAW04644;
 XX
 XX 01-AUG-1997 (first entry)
 XX
 XX Ovalbumin-derived activated CD8+ T cells epitope OVA10C.
 DE
 XX Macrophage; artificial antigen presenting cell; APC; cancer;
 KW tumours; neoplasia; viral infection; retroviral infection;
 KW autoimmune.
 XX
 XX Synthetic.
 OS
 XX WO9637107-A1.
 XX
 XX 28-NOV-1996.
 XX

```

PF 22-MAY-1996; 96WO-US07436.
XX
PR 23-MAY-1995; 95US-0447761.
XX
PA (SCRI ) SCRIPPS RES INST.
XX
PI DeBruijn MLH, Jackson MR, Peterson PA;
XX
DR WPI; 1997-020850/02.
XX
XX Prodn. of activated CD8+ T cells directed to specific antigen - can
PT specifically kill target cells useful to treat, e.g. cancer
PT
PS Example 1; Page 26; 84pp; English.
XX
XX The method for the production of activated CD8+ T cells specifically
CC directed towards a particular antigen involves affixing peptides
CC corresponding to the particular antigen to an artificial support;
CC contacting macrophages with the affixed peptides for a time sufficient
CC for the peptides to be engulfed, and at least a portion of the peptides
CC to be presented on the surface of the macrophage; and contacting
CC unprimed CD8+ T cells with the peptide presenting macrophages for a
CC time sufficient to activate the unprimed CD8+ T cells. The present
CC sequence represents a peptide designated OVA10C which corresponds to
CC ovalbumin, a Kb-restricted peptide antigen. This represents the optimal
CC peptide with the addition of two amino acids at the carboxy-terminus.
CC Small extensions to the optimal peptide affect the affinity of the
CC peptide for soluble class I molecules in vitro e.g. the addition
CC of two amino acids to the amino-terminus lowers the affinity to Kb by
CC 76-fold compared to the optimal peptide; addition of two amino acids to
CC the carboxy-terminus lowers the affinity by 4-fold. The method,
CC macrophages and artificial antigen presenting cell, having a peptide
CC corresponding to the particular antigen present on its surface and at
CC least a portion of an artificial support in its interior, can be used to
CC treat conditions (e.g. cancer, tumours, neoplasia, viral or retroviral
CC infection or autoimmune or autoimmune-type conditions) in patients via
CC the specific killing of target cells.
XX
SQ Sequence 10 AA;
Query Match 100.0%; Score 38; DB 18; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
Db |||||
1 SIINFEKL 8

RESULT 78
AAU09821
ID AAU09821 standard; peptide; 10 AA.
XX
AC AAU09821;
XX
XX 14-FEB-2002 (first entry)
XX
DE Modified ovalbumin-derived class I H-2Kb restricted peptide #1.
XX
XX Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW popliteal lymph node; spleen; immune response; systemic response.
XX
OS Synthetic.
XX
XX WO200178767-A2.
XX
XX 25-OCT-2001.
XX
XX 17-APR-2001; 2001WO-EP04313.
XX
XX 14-APR-2000; 2000AT-0000657.
XX
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX Mattner F, Zauner W, Schmidt W, Buschle M;
XX
XX WPI; 2002-025970/03.
XX
XX Pharmaceutical preparation for use as a potent vaccine for inducing an
PT improved immune response in a mammal, comprises a modified peptide -
PT
XX
XX Example 1; Page 9; 18pp; English.
XX
XX The invention relates to a pharmaceutical preparation comprising a
CC modified peptide, which induces an improved immune response in a mammal
CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
CC negatively charged (Glu), one positively charged (Lys) amino acid) was
CC rendered negative by adding (at the N-terminus) Glu or Glu Asp Glu
CC Asp, respectively. Results showed that the addition of 4 negatively-
CC charged amino acids (EDSD) at the N-terminus of peptide SIINFEKL makes
CC this peptide (in combination with poly-L-arginine) able to induce a high
CC amount of specific interferon (IFN)-gamma-producing T cells in the
CC draining (popliteal) lymph node (local response) and in the spleen
CC (systemic response). Thus, the addition of hydrophobic amino acids as
CC well as the addition of negatively charged amino acids transforms the
CC peptide SIINFEKL to a good inducer of specific T cells. The modified
CC peptides of the pharmaceutical composition induce a stronger immune
CC response in a mammal compared to wild type antigens. The present
CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
CC peptide #1 as described in the method of the invention.
XX
SQ Sequence 10 AA;
Query Match 100.0%; Score 38; DB 23; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
Db |||||
3 SIINFEKL 10

RESULT 79
AAU09825
ID AAU09825 standard; peptide; 10 AA.
XX
AC AAU09825;
XX
XX 14-FEB-2002 (first entry)
XX
DE Modified ovalbumin-derived class I H-2Kb restricted peptide #5.
XX
XX Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW popliteal lymph node; spleen; immune response; systemic response.
XX
OS Synthetic.
XX
XX WO200178767-A2.
XX
XX 25-OCT-2001.
XX
XX 17-APR-2001; 2001WO-EP04313.
XX
XX 14-APR-2000; 2000AT-0000657.
XX
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX Mattner F, Zauner W, Schmidt W, Buschle M;
XX
XX WPI; 2002-025970/03.
XX
XX Pharmaceutical preparation for use as a potent vaccine for inducing an
PT improved immune response in a mammal, comprises a modified peptide -
PT
XX

```

```

Db      331 SIINFEXL 338

RESULT 125
AAB31611
ID      AAB31611 standard; Protein; 948 AA.
XX
AC      AAB31611;
XX
DT      30-APR-2001 (first entry)
XX
DE      Amino acid sequence of Hsp65-ovalbumin fusion protein.
XX
KW      Heat shock protein; Hsp; Th1 response; Th1 cell; CD4+ T lymphocyte cell;
KW      lymphocyte; Hsp65; Hsp40; Hsp10; Hsp60; Hsp71; microbial pathogen;
KW      ovalbumin.
XX
OS      Synthetic.
OS      Mycobacterium bovis.
OS      Gallus sp.
XX
PN      WO200104344-A2.
XX
PD      18-JAN-2001.
XX
PF      10-JUL-2000; 2000WO-US18828.
XX
PR      08-JUL-1999; 99US-0143757.
XX
PA      (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI      Siegel M, Chu NR, Mizzen LA;
XX
DR      WPI; 2001-138361/14.
DR      N-PSDE; AAF25014.
XX
PT      Screening for compounds that stimulate Th1-like responses in CD4+ T
PT      lymphocyte cells -
XX
PS      Example 8; Fig 7A-B; 88pp; English.
XX
CC      The present sequence represents a fusion protein comprising a
CC      Mycobacterium bovis heat shock protein (Hsp) 65 fused at its C terminal
CC      to an ovalbumin protein. The fusion protein is used in the method of the
CC      invention. The specification describes a method of determining whether a
CC      compound stimulates a Th1-like response. Th1 cells are a subset of CD4+
CC      T lymphocyte cells. The method comprises contacting naive lymphocytes
CC      in vitro with a fusion protein comprising at least a fragment of Hsp,
CC      and then detecting the Th1-like response exhibited by the cell sample.
CC      The proteins which may be used in the method of the invention are Hsp65,
CC      Hsp40, Hsp10, Hsp60, and Hsp71. The method may be used to identify
CC      compounds that stimulate Th1-like responses in response to microbial
CC      pathogens.
XX
SQ      Sequence 948 AA;

Query Match      100.0%; Score 38; DB 22; Length 948;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SIINFEXL 8
        |||||
Db      820 SIINFEXL 827

Search completed: January 30, 2004, 07:17:23
Job time : 74 secs

```

```

CC Yeast/mouse fusion gene, as described above.
XX
SQ Sequence 541 AA;
    Query Match      100.0%; Score 38; DB 23; Length 541;
    Best Local Similarity 100.0%; Pred. No. 9.7; 0; Indels 0; Gaps 0;
    Matches 8; Conservative 0; Mismatches 0;

QY 1 SIINFEKL 8
Db 34 SIINFEKL 41
    |||||
    |||||

RESULT 123
ID AAE13110 standard; Protein; 564 AA.
XX AC AAE13110;
XX DT 28-JAN-2002 (first entry)
DE Human HER500 fusion protein construct comprising OVA-derived octapeptide.
XX KW Immunostimulatory fusion protein; IFP; antigen component; therapy;
KW immunostimulatory component; T-cell mediated immune response; DC;
KW dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
KW PAP protein; Ala Arg linker; membrane distal extracellular domain;
KW membrane distal intracellular domain; C-terminal tag; human; OVA;
KW HER-2 protein; ovalbumin-derived octapeptide; HER500 fusion protein.
XX OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX PN WO200174855-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US10515.
XX PR 30-MAR-2000; 2000US-193504P.
XX PA (DEND-) DENDREON CORP.
XX PI Laus R, Vidovic D, Graddis T;
XX WPI; 2001-662965/76.
XX DR N-PSDB; AAD21566.
XX PS Claim 7; Page 26; 59pp; English.
XX CC The invention relates to immunostimulatory fusion proteins (IFP) and
CC nucleic acid molecules encoding such proteins. The IFPs comprise a
CC polypeptide antigen component and an immunostimulatory component derived
CC from the intracellular domain of HER-2 protein which is effective to
CC elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
CC immune response to the antigen. IFP or superactivated dendritic cells
CC are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
CC associated with a particularly antigen. The present sequence is HER500
CC fusion protein construct which comprises human PAP signal
CC sequence, mature PAP protein, an Ala Arg linker, human HER-2 signal
CC sequence, mature HER-2 membrane distal extracellular domain, an
CC Ala linker, an ovalbumin (OVA)-derived immunodominant octapeptide,
CC HER-2 membrane distal intracellular domain and a C-terminal tag.
XX SQ Sequence 564 AA;
    Query Match      100.0%; Score 38; DB 22; Length 564;
    Best Local Similarity 100.0%; Pred. No. 10;
    Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC Yeast/mouse fusion gene, as described above.
XX
SQ Sequence 541 AA;
    Query Match      100.0%; Score 38; DB 23; Length 541;
    Best Local Similarity 100.0%; Pred. No. 9.7; 0; Indels 0; Gaps 0;
    Matches 8; Conservative 0; Mismatches 0;

QY 1 SIINFEKL 8
Db 331 SIINFEKL 338
    |||||
    |||||

RESULT 124
ID AAE13111 standard; Protein; 697 AA.
XX AC AAE13111;
XX DT 28-JAN-2002 (first entry)
DE Human HER500-rGM-CSF fusion construct comprising OVA-derived peptide.
XX KW Immunostimulatory fusion protein; IFP; antigen component; therapy;
KW immunostimulatory component; T-cell mediated immune response; DC;
KW dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
KW PAP protein; Ala Arg linker; membrane distal extracellular domain;
KW membrane distal intracellular domain; C-terminal tag; human; GM-CSF;
KW HER-2 protein; granulocyte-macrophage colony stimulating factor;
KW ovalbumin-derived octapeptide; OVA; rat; HER500-rGM-CSF fusion protein.
XX OS Chimeric - Homo sapiens.
OS Chimeric - Rattus norvegicus.
XX OS Chimeric - Unidentified.
XX PN WO200174855-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US10515.
XX PR 30-MAR-2000; 2000US-193504P.
XX PA (DEND-) DENDREON CORP.
XX PI Laus R, Vidovic D, Graddis T;
XX WPI; 2001-662965/76.
XX DR N-PSDB; AAD21567.
XX PS Claim 7; Page 27; 59pp; English.
XX CC The invention relates to immunostimulatory fusion proteins (IFP) and
XX nucleic acid molecules encoding such proteins. The IFPs comprise a
XX polypeptide antigen component and an immunostimulatory component derived
XX from the intracellular domain of HER-2 protein which is effective to
XX elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
XX immune response to the antigen. IFP or superactivated dendritic cells
XX are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
XX associated with a particularly antigen. The present sequence is HER500
XX rGM-CSF fusion protein construct which comprises human PAP
XX signal sequence, mature PAP protein, an Ala Arg linker, human HER-2
XX signal sequence, mature HER-2 membrane distal extracellular domain,
XX an Ala linker, an ovalbumin (OVA)-derived immunodominant octapeptide,
XX HER-2 membrane distal intracellular domain, an Ala linker, a mature
XX rat granulocyte-macrophage colony stimulating factor (GM-CSF) sequence
XX and a C-terminal tag.
XX SQ Sequence 697 AA;
    Query Match      100.0%; Score 38; DB 22; Length 697;
    Best Local Similarity 100.0%; Pred. No. 13;
    Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
    |||||
    |||||
  
```

CC used to identify compounds that stimulate Th1-like responses in response
 CC to microbial pathogens.

SQ Sequence 409 AA;
 Query Match 100.0%; Score 38; DB 22; Length 409;
 Best Local Similarity 100.0%; Pred. No. 7.2; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 281 SIINFEKL 288

RESULT 121
 AAE13112
 ID AAE13112 standard; Protein; 479 AA.
 AC AAE13112;
 DT 28-JAN-2002 (first entry)

XX Human HER300-RGM-CSF fusion construct comprising OVA-derived peptide.
 DE Immunostimulatory fusion protein; IFP; antigen component; therapy;
 KW immunostimulatory component; T-cell mediated immune response; DC;
 KW dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
 KW PAP protein; Ala Arg linker; membrane distal extracellular domain;
 KW membrane distal intracellular domain; C-terminal tag; human; GM-CSF;
 KW HER-2 protein; granulocyte-macrophage colony stimulating factor;
 KW ovalbumin-derived octapeptide; OVA; rat; HER300-RGM-CSF fusion protein.

XX Chimeric - Homo sapiens.
 OS Chimeric - Rattus norvegicus.
 OS Chimeric - Unidentified.

XX WO200174855-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US10515.
 XX 30-MAR-2000; 2000US-193504P.
 XX (DEND-) DENDREON CORP.

PI Laus R, Vidovic D, Graddis T;
 XX WPI; 2001-662965/76.
 XX N-PSDB; AAD21568.

PT An immunostimulatory fusion protein comprising the intracellular domain
 of HER-2 and an antigen elicits an immune response to the antigen and
 is useful for the treatment of associated cancer associated -
 XX Example 1; Page 27; 59pp; English.

XX The invention relates to immunostimulatory fusion proteins (IFP) and
 nucleic acid molecules encoding such proteins. The IFPs comprise a
 polypeptide antigen component and an immunostimulatory component derived
 from the intracellular domain of HER-2 protein which is effective to
 elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
 immune response to the antigen. IFP or superactivated dendritic cells
 are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
 associated with a particularly antigen. The present sequence is HER300
 RGM-CSF fusion protein construct which comprises human PAP
 signal sequence, mature PAP protein, an Ala Arg linker, human HER-2
 an Ala linker, an ovalbumin (OVA)-derived immunodominant octapeptide,
 an Ala linker, a HER-2 membrane distal intracellular domain, a mature
 rat granulocyte-macrophage colony stimulating factor (GM-CSF) sequence
 and a C-terminal tag.

SQ Sequence 479 AA;

Query Match 100.0%; Score 38; DB 22; Length 479;
 Best Local Similarity 100.0%; Pred. No. 8.5; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 |||||
 DB 331 SIINFEKL 338

RESULT 122
 AAU99725
 ID AAU99725 standard; Protein; 541 AA.
 AC AAU99725;
 DT 07-OCT-2002 (first entry)

XX Yeast/mouse SS-OVA-Kb/beta2m-c-myc-AGA2 fusion protein.
 DE Mutant major histocompatibility complex class I chimeric protein; MHC;
 KW lymphocyte; T-cell receptor; tissue sample; biopsy material; pathogen;
 KW bodily fluid; T lymphocyte; neoplastic cell; tumour cell; MHC antigen;
 KW virus; protozoan; bacteria; fungi; nematode; immune response; activator;
 KW enhancer; T cell activator; mouse; recombinant yeast cell; Kb; OVA;
 KW beta2m; dEV8; AGA2; SIYK; fusion protein.

XX Chimeric - Mus sp.
 OS Chimeric - Saccharomyces cerevisiae.
 OS Synthetic.

XX WO200246399-A2.
 XX 13-JUN-2002.
 XX 10-DEC-2001; 2001WO-US47817.
 XX 08-DEC-2000; 2000US-254495P.
 XX (UNII) UNIV ILLINOIS FOUND.

PI Kranz DM, Brophy S;
 XX WPI; 2002-527916/56.
 XX N-PSDB; ABR87870.

PT New isolated mutant major histocompatibility complex class I chimeric
 protein displayed on surfaces of recombinant yeast cells, has improved
 stability, and is useful for activating immune response -
 XX Example 7; Page 38-39; 96pp; English.

XX The present invention relates to a new mutant major histocompatibility
 complex (MHC) class I chimeric protein. The protein of the invention
 comprises a portion mediating binding to surfaces of recombinant yeast
 cells and a portion comprising peptide binding region of MHC class I
 protein, where the invention is improved in stability as compared with
 MHC class I chimeric protein which is not a mutant chimeric protein.
 The protein, further comprising a detectable label, is useful for
 detecting a lymphocyte having a T-cell receptor protein in a biological
 sample such as cells, tissue sample, biopsy material or bodily fluids.
 The method is useful for detecting a T lymphocyte that is specific for
 a neoplastic cell, a tumour cell, a virus-infected cell, a protozoan-
 infected cell, a bacterium-infected cell or a fungus-infected cell. The
 protein of the invention can be used to directly activate T cells, in
 order to identify/screen for peptide-MHC antigens. The protein is also
 useful in activating T cells that participate in the removal of target
 cells including neoplastic cells and cells infected with pathogenic
 agents including viruses, protozoans, bacteria, fungi or nematodes.
 The invention is improved in stability as compared with MHC class I
 protein which is not a mutant chimeric protein. The present amino acid
 sequence represents a chimeric MHC protein that is encoded by a

CC in an exemplification of the present invention. The invention
 CC relates to a novel expression vector comprising a promoter operably
 CC linked to a fusion gene encoding a major histocompatibility complex
 CC (MHC) targeting sequence, and two or more heterologous peptide
 CC epitopes. The MHC targeting sequence may be a class I targeting
 CC sequence, which directs an MHC class I epitope to a cytosolic pathway or
 CC to the endoplasmic reticulum, or an MHC class II targeting sequence,
 CC which directs extracellular antigens to enter the endocytic pathway to
 CC be processed into antigen peptides for presentation on MHC class II
 CC molecules. The heterologous epitopes may comprise either helper T
 CC lymphocyte (HTL) epitopes, or a cytotoxic T lymphocyte (CTL) epitope and
 CC a universal HTL epitope such as a pan DR epitope (PADRE). The vectors
 CC are useful for stimulating an immune response in vivo, as well as for use
 CC in assaying the human immunogenicity of a human T cell peptide epitope
 CC in vivo in a non-human mammal. They provide a nucleic acid vaccine for
 CC enhancing immunity against infectious pathogens, such as viruses (e.g.,
 CC HIV, hepatitis B (HBV) and hepatitis C (HCV)), bacteria, protozoa (e.g.,
 CC Plasmodium falciparum, the cause of malaria) and also tumour cells and
 CC autoimmune diseases. Universal MHC class I and class II epitopes are advantageously
 CC combined with other MHC class I and class II epitopes to increase the
 CC number of cells that are activated in response to a given antigen and
 CC provide a broader population coverage of MHC-reactive alleles.

XX SQ Sequence 132 AA;
 Query Match 100.0%; Score 38; DB 21; Length 132;
 Best Local Similarity 100.0%; Pred. No. 2.2; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0;
 QY 1 SIINFELK 8
 DB 75 SIINFELK 82

RESULT 119
 AAE13435
 ID AAE13435 standard; Protein; 386 AA.
 AC AAE13435;
 XX 12-FEB-2002 (first entry)
 DT Chicken ovalbumin containing plurality of epitopes.
 DE Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
 KW major histocompatibility complex; MHC; therapy; immune response;
 KW malignancy; chicken.
 XX Gallus gallus.
 OS
 XX Key Location/Qualifiers
 FH Domain 200..291
 FT Region 258..265
 FT /note= "MHC class I epitope"
 FT Region 266..281
 FT /note= "MHC class II epitope"
 XX WO200179259-A1.
 PN 25-OCT-2001.
 XX 17-APR-2001; 2001WO-US12567.
 XX 17-APR-2000; 2000US-197462P.
 XX (ROTH/) ROTHMAN J E.
 PA (MAYH/) MAYHEW M.
 PA (HOEM/) HOE M.
 XX Rothman JE, Mayhew M, Hoe M;
 XX WPI; 2002-017594/02.
 DR N-PSDB; AAD22407.

XX A new antigenic complex comprising epitopes non-covalently joined to a
 PT heat shock protein by a molecular tether designated a javelin are
 PT useful to treat or prevent infectious disease or malignancy -
 PS Example; Fig 2; 47pp; English.
 XX The present invention relates to an antigenic complex, comprising a
 CC number of epitopes non-covalently joined to a heat shock protein (HSP) by
 CC a tethering molecule referred to as javelin which has affinity for the
 CC HSP under physiological conditions, where the epitopes are covalently
 CC joined to the tethering molecule and one epitope is major
 CC histocompatibility complex class I (MHC) and the other MHC class II. The
 CC antigenic complex is used to induce immune responses directed towards the
 CC treatment or prevention of infectious diseases and malignancies. The
 CC present sequence is chicken ovalbumin containing plurality of epitopes.
 XX SQ Sequence 386 AA;
 Query Match 100.0%; Score 38; DB 23; Length 386;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFELK 8
 DB 258 SIINFELK 265
 RESULT 120
 AAB31545
 ID AAB31545 standard; Protein; 409 AA.
 XX AAB31545;
 AC AAB31545;
 XX 30-APR-2001 (first entry)
 DT Amino acid sequence of chicken ovalbumin.
 DE Heat shock protein; Hsp; Th1 response; Th1 cell; CD4+ T lymphocyte cell;
 KW lymphocyte; Hsp65; Hsp40; Hsp10; Hsp60; Hsp71; microbial pathogen;
 KW ovalbumin.
 XX Gallus sp.
 OS
 XX WO200104344-A2.
 PN 18-JAN-2001.
 XX 10-JUL-2000; 2000WO-US18828.
 XX 08-JUL-1999; 99US-0143757.
 XX (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
 XX Siegel M, Chu NR, Mizzen LA;
 XX WPI; 2001-138361/14.
 DR N-PSDB; AAF25127.
 XX Screening for compounds that stimulate Th1-like responses in CD4+ T
 PT lymphocyte cells -
 XX Example 8; Fig 6; 88pp; English.
 XX The present sequence represents an ovalbumin protein. Ovalbumin was fused
 CC to a heat shock protein (Hsp), and used in the method of the
 CC invention. The specification describes a method of determining whether
 CC a compound stimulates a Th1-like response. Th1 cells are a subset of
 CC CD4+ T lymphocyte cells. The method comprises contacting naive
 CC lymphocytes in vitro with a fusion protein comprising at least a
 CC fragment of Hsp, and then detecting the Th1-like response exhibited by
 CC the cell sample. The proteins which may be used in the method of the
 CC invention are Hsp65, Hsp40, Hsp10, Hsp60, and Hsp71. The method may be

PT heat shock protein by a molecular tether designated a javelin are
 PT useful to treat or prevent infectious disease or malignancy -
 XX
 XX Disclosure; Page 15; 47pp; English.
 XX
 CC The present invention relates to an antigenic complex, comprising a
 CC number of epitopes non-covalently joined to a heat shock protein (HSP) by
 CC a tethering molecule referred to as javelin which has affinity for the
 CC HSP under physiological conditions, where the epitopes are covalently
 CC joined to the tethering molecule and one epitope is major
 CC histocompatibility complex class I (MHC) and the other MHC class II. The
 CC antigenic complex is used to induce immune responses directed towards the
 CC treatment or prevention of infectious diseases and malignancies. The
 CC present sequence is chicken ovalbumin derived protein domain.
 XX
 XX Sequence 108 AA;
 QY Query Match 100.0%; Score 38; DB 23; Length 108;
 DB Best Local Similarity 100.0%; Pred. No. 1.8; 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB 66 SIINFEKL 73
 RESULT 117
 AAEL13463
 ID AAE13463 standard; Protein; 111 AA.
 AC AAE13463;
 XX
 DT 12-FEB-2002 (first entry)
 DE Chicken ovalbumin derived protein domain #6.
 XX
 KW Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
 KW major histocompatibility complex; MHC; therapy; immune response;
 KW malignancy; chicken.
 XX
 OS Gallus gallus.
 FH Key Location/Qualifiers
 FT Region 1..8
 FT /note= "Javelin sequence"
 FT Region 9..11
 FT /note= "Linker"
 FT Region 69..76
 FT /note= "MHC class I epitope"
 FT Region 77..92
 FT /note= "MHC class II epitope"
 FT Region 104..111
 FT /note= "Javelin sequence"
 WO200179259-A1.
 XX
 XX 25-OCT-2001.
 XX
 XX 17-APR-2001; 2001WO-US12567.
 XX
 XX 17-APR-2000; 2000US-197462P.
 XX
 PA (ROTH/) ROTHMAN J E.
 PA (MAYH/) MAYHEW M.
 PA (HOEM/) HOE M.
 XX
 PI Rothman JB, Mayhew M, Hoe M;
 XX
 DR WPI; 2002-017594/02.
 XX
 CC A new antigenic complex comprising epitopes non-covalently joined to a
 PT heat shock protein by a molecular tether designated a javelin are
 PT useful to treat or prevent infectious disease or malignancy -

XX
 PS Disclosure; Page 15; 47pp; English.
 XX
 CC The present invention relates to an antigenic complex, comprising a
 CC number of epitopes non-covalently joined to a heat shock protein (HSP) by
 CC a tethering molecule referred to as javelin which has affinity for the
 CC HSP under physiological conditions, where the epitopes are covalently
 CC joined to the tethering molecule and one epitope is major
 CC histocompatibility complex class I (MHC) and the other MHC class II. The
 CC antigenic complex is used to induce immune responses directed towards the
 CC treatment or prevention of infectious diseases and malignancies. The
 CC present sequence is chicken ovalbumin derived protein domain.
 XX
 XX Sequence 111 AA;
 QY Query Match 100.0%; Score 38; DB 23; Length 111;
 DB Best Local Similarity 100.0%; Pred. No. 1.9; 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB 69 SIINFEKL 76
 RESULT 118
 AAY52575
 ID AAY52575 standard; Protein; 132 AA.
 AC AAY52575;
 XX
 DT 28-FEB-2000 (first entry)
 DE Amino acid sequence of AOS minigene insert of expression vector pMIN.0.
 XX
 KW Chimeric; pan DR epitope; expression vector;
 KW promoter; major histocompatibility complex; MHC; targeting; peptide;
 KW epitope; antigen; presentation; class II; cytosolic pathway;
 KW endoplasmic reticulum; class II; extracellular antigen;
 KW endocytic pathway; helper T lymphocyte; HTL; universal epitope;
 KW cytotoxic T lymphocyte; CTL; immune response; immunogenicity; assay;
 KW vaccine; immunity; infection; pathogen; virus; HIV; HBV; HCV;
 KW hepatitis B; hepatitis C; bacterium; protozoan; tumour cell;
 KW autoimmune disease; activation; antiviral; antimalarial;
 KW immunoprotective; minigene.
 XX
 OS Synthetic.
 XX
 XX WO958658-A2.
 XX
 PD 18-NOV-1999.
 XX
 XX 13-MAY-1999; 99WO-US10646.
 XX
 XX 13-MAY-1998; 98US-0078904.
 XX
 XX 15-MAY-1998; 98US-0085751.
 XX
 PA (EPIM-) EPIMUNE INC.
 XX
 XX Fikes JD, Hermanson GG, Sette A, Ishioka GY, Livingston B;
 XX Chesnut RW;
 XX
 DR WPI; 2000-039103/03.
 DR N-PSDB; AAZ38634.
 XX
 PT Expression vectors encoding major histocompatibility targeting
 PT sequence, used as, e.g. tumor vaccines -
 XX
 XX Example 1; Fig 20; 130pp; English.
 XX
 CC This sequence represents the amino acid sequence of the AOS minigene
 CC insert of the expression vector pMIN.0 (AAZ38634). This insert encodes
 CC several MHC class I epitopes, and also the universal MHC
 CC class II (helper T) epitope, pan DR epitope (PADRE), and was used

KW Polypeptide; vaccinia; cytotoxic T lymphocyte; CTL; mouse; epitope; vaccine;
 KW major histocompatibility complex; pathogen; HLA diversity; avipox virus;
 KW bacterial vector; rhabdovirus vector; ISCOM; influenza nuclear protein;
 KW ovalbumin; cytomagalovirus; adenovirus; sendai virus; P.Berghei; MHC;
 KW circumsporozoite protein; influenza NS1; choriomeningitis virus.
 XX Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Region 4..12 /note= "cytotoxic T lymphocyte (CTL) epitope #1, isolated
 FT from influenza nuclear protein (residues
 FT 366-374)"
 FT Region 13..20 /note= "CTL epitope #2, isolated from ovalbumin (residues
 FT 257-264)"
 FT Region 21..29 /note= "CTL epitope #3, isolated from influenza nuclear
 FT protein (residues 147-155)"
 FT Region 30..37 /note= "CTL epitope #4, isolated from influenza nuclear
 FT protein (residues 50-58)"
 FT Region 38..46 /note= "CTL epitope #5, isolated from murine
 FT cytomagalovirus pp89 (residues 168-176)"
 FT Region 48..58 /note= "CTL epitope #6, isolated from adenovirus 5 E1A
 FT (residues 234-243)"
 FT Region 59..67 /note= "CTL epitope #7, isolated from sendai virus
 FT nuclear protein (residues 324-332)"
 FT Region 68..76 /note= "CTL epitope #8, isolated from P.Berghei
 FT circumsporozoite protein (residues 249-257)"
 FT Region 77..85 /note= "CTL epitope #9, isolated from influenza NS1
 FT (residues 152-160)"
 FT Misc-difference 78 /note= "encoded by GAC"
 FT Region 86..94 /note= "CTL epitope #10, isolated from lymphocytic
 FT choriomeningitis virus nuclear protein (residues
 FT 118-126)"
 FT Region 97..106 /note= "monoclonal antibody epitope"
 FT Misc-difference 105 /note= "encoded by AGA"
 FT
 XX WO9603144-A1.
 XX 08-FEB-1996.
 XX 27-JUL-1995; 95WO-AU00461.
 XX 08-FEB-1995; 95AU-0001009.
 XX 27-JUL-1994; 94AU-0007079.
 XX (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
 XX (CSIR) COMMONWEALTH SCI & IND RES ORG.
 XX (UTME) UNIV MELBOURNE.
 XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX (BIOT-) BIOTECH AUSTRALIA PTY LTD.
 XX (CSLC-) CSL LTD.
 XX Burrows SR, Coupar BEH, Khanna R, Moss DJ, Suhrbier A;
 FI Thomson SA;
 XX WPI; 1996-116788/12.
 DR N-PSDB; AAT12413.
 XX New poly:epitope cytotoxic T lymphocyte vaccines - comprising a
 FT recombinant protein including CTL epitope(s) from pathogens, free of
 FT natural flanking sequences

XX Claim 5; Fig 5; 46pp; English.
 PS This sequence represents a polytope encoded by a DNA insert of a
 CC recombinant vaccinia virus of the invention. This sequence contains 10
 CC murine cytotoxic T lymphocyte (CTL) epitopes. Each of the epitopes in
 CC this sequence are capable of producing a primary CTL response in mice
 CC with the appropriate major histocompatibility complex (MHC) allele. This
 CC sequence (and the DNA encoding it) can be used in vaccines against
 CC multiple epitopes derived from several different pathogens. The vaccine
 CC could alternatively contain a large number of epitopes from one pathogen
 CC so that HLA diversity of the target population is covered. The vaccines
 CC can be delivered by vaccinia virus, avipox virus, bacterial or
 CC rhabdovirus vectors, or by virus-like particles. The proteins are
 CC preferably administered with ISCOMs when they are delivered directly.
 CC The advantage with these vaccines is that they provide a more diverse
 CC immune response.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX SQ Sequence 106 AA;
 Query Match 100.0%; Score 38; DB 17; Length 106;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 SIINFEXL 8
 Db 13 SIINFEXL 20
 RESULT 116
 AAE13462
 ID AAE13462 standard; Protein; 108 AA.
 XX AC AAE13462;
 XX DT 12-FEB-2002 (first entry)
 XX DE Chicken ovalbumin derived protein domain #5.
 XX KW Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
 KW major histocompatibility complex; MHC; therapy; immune response;
 KW malignancy; chicken.
 XX OS Gallus gallus.
 XX FH Key Location/Qualifiers
 FT Region 1..8
 FT /note= "Javelin sequence"
 FT Region 56..73
 FT /note= "MHC class I epitope"
 FT Region 74..89
 FT /note= "MHC class II epitope"
 FT Region 101..108
 FT /note= "Javelin sequence"
 XX WO200179259-A1.
 XX 25-OCT-2001.
 XX 17-APR-2001; 2001WO-US12567.
 XX 17-APR-2000; 2000US-197462P.
 XX (ROTH/) ROTHMAN J E.
 XX (MAYH/) MAYHEW M.
 XX (HOEM/) HOE M.
 XX Rothman JE, Mayhew M, Hoe M;
 XX WPI; 2002-017594/02.
 XX A new antigenic complex comprising epitopes non-covalently joined to a

```

XX AC AAE13459;
XX DT 12-FEB-2002 (first entry)
XX DE Chicken ovalbumin derived protein domain #2.
XX KW Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
XX KW major histocompatibility complex; MHC; therapy; immune response;
XX OS Gallus gallus.
XX FH Key Location/Qualifiers
XX FT Region 1..8 /note= "Javelin sequence"
XX FT Region 9..11 /note= "Linker"
XX FT Region 69..76 /note= "MHC class I epitope"
XX FT Region 77..92 /note= "MHC class II epitope"
XX PN WO200179259-A1.
XX PD 25-OCT-2001.
XX PF 17-APR-2001; 2001WO-US12567.
XX PR 17-APR-2000; 2000US-197462P.
XX PA (ROTH/) ROTHMAN J E.
XX PA (MAYH/) MAYHEW M.
XX PA (HOEM/) HOE M.
XX PI Rothman JE, Mayhew M, Hoe M;
XX WPI; 2002-017594/02.
XX A new antigenic complex comprising epitopes non-covalently joined to a
XX heat shock protein by a molecular tether designated a javelin are
XX useful to treat or prevent infectious disease or malignancy -
XX Disclosure; Page 14; 47pp; English.
XX The present invention relates to an antigenic complex, comprising a
XX number of epitopes non-covalently joined to a heat shock protein (HSP) by
XX a tethering molecule referred to as javelin which has affinity for the
XX HSP under physiological conditions, where the epitopes are covalently
XX joined to the tethering molecule and one epitope is major
XX histocompatibility complex class I (MHC) and the other MHC class II. The
XX antigenic complex is used to induce immune responses directed towards the
XX treatment or prevention of infectious diseases and malignancies. The
XX present sequence is chicken ovalbumin derived protein domain.
XX SQ Sequence 103 AA;
XX Query Match 100.0%; Score 38; DB 23; Length 103;
XX Best Local Similarity 100.0%; Pred. NO. 1.7;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SIINFEKL 8
XX |||||
XX 69 SIINFEKL 76
XX
XX RESULT 114
XX AAE13461
XX ID AAE13461 standard; Protein; 103 AA.
XX AC AAE13461;
XX DT 12-FEB-2002 (first entry)

```

```

XX DE Chicken ovalbumin derived protein domain #4.
XX KW Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
XX KW major histocompatibility complex; MHC; therapy; immune response;
XX OS Gallus gallus.
XX FH Key Location/Qualifiers
XX FT Region 58..65 /note= "MHC class I epitope"
XX FT Region 66..81 /note= "MHC class II epitope"
XX FT Region 93..95 /note= "Linker sequence"
XX FT Region 96..103 /note= "Javelin sequence"
XX PN WO200179259-A1.
XX PD 25-OCT-2001.
XX PF 17-APR-2001; 2001WO-US12567.
XX PR 17-APR-2000; 2000US-197462P.
XX PA (ROTH/) ROTHMAN J E.
XX PA (MAYH/) MAYHEW M.
XX PA (HOEM/) HOE M.
XX PI Rothman JE, Mayhew M, Hoe M;
XX WPI; 2002-017594/02.
XX A new antigenic complex comprising epitopes non-covalently joined to a
XX heat shock protein by a molecular tether designated a javelin are
XX useful to treat or prevent infectious disease or malignancy -
XX Disclosure; Page 14; 47pp; English.
XX The present invention relates to an antigenic complex, comprising a
XX number of epitopes non-covalently joined to a heat shock protein (HSP) by
XX a tethering molecule referred to as javelin which has affinity for the
XX HSP under physiological conditions, where the epitopes are covalently
XX joined to the tethering molecule and one epitope is major
XX histocompatibility complex class I (MHC) and the other MHC class II. The
XX antigenic complex is used to induce immune responses directed towards the
XX treatment or prevention of infectious diseases and malignancies. The
XX present sequence is chicken ovalbumin derived protein domain.
XX SQ Sequence 103 AA;
XX Query Match 100.0%; Score 38; DB 23; Length 103;
XX Best Local Similarity 100.0%; Pred. NO. 1.7;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SIINFEKL 8
XX |||||
XX 58 SIINFEKL 65
XX
XX RESULT 115
XX AAR89966
XX ID AAR89966 standard; Protein; 106 AA.
XX AC AAR89966;
XX DT 25-MAR-2003 (updated)
XX DT 12-SEP-1996 (first entry)
XX DE Polypeptide sequence.
XX

```

Query Match 100.0%; Score 38; DB 22; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.93;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
|||||
DB 10 SIINFEKL 17

RESULT 111
AAE13458
ID AAE13458 standard; Protein; 100 AA.

XX

AC AAE13458;

XX 12-FEB-2002 (first entry)

XX Chicken ovalbumin derived protein domain #1.

XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;

XX major histocompatibility complex; MHC; therapy; immune response;

XX malignancy; chicken.

XX Gallus gallus.

XX Key Location/Qualifiers

XX 1..8 /note= "Javelin sequence"

XX 66..73 /note= "MHC class I epitope"

XX 74..89 /note= "MHC class II epitope"

XX WO200179259-A1.

XX 25-OCT-2001.

XX 17-APR-2001; 2001WO-US12567.

XX 17-APR-2000; 2000US-197462P.

XX (ROTH/) ROTHMAN J E.

XX (MAYH/) MAYHEW M.

XX (HOEW/) HOE M.

XX Rothman JE, Mayhew M, Hoe M;

XX WPI; 2002-017594/02.

XX A new antigenic complex comprising epitopes non-covalently joined to a

XX heat shock protein by a molecular tether designated a javelin are

XX useful to treat or prevent infectious disease or malignancy -

XX Disclosure; Page 14; 47pp; English.

XX The present invention relates to an antigenic complex, comprising a

XX number of epitopes non-covalently joined to a heat shock protein (HSP) by

XX a tethering molecule referred to as javelin which has affinity for the

XX HSP under physiological conditions, where the epitopes are covalently

XX joined to the tethering molecule and one epitope is major

XX histocompatibility complex class I (MHC) and the other MHC class II. The

XX antigenic complex is used to induce immune responses directed towards the

XX treatment or prevention of infectious diseases and malignancies. The

XX present sequence is chicken ovalbumin derived protein domain.

DB 66 SIINFEKL 73

RESULT 112

AAE13460

ID AAE13460 standard; Protein; 100 AA.

XX

AC AAE13460;

XX 12-FEB-2002 (first entry)

XX Chicken ovalbumin derived protein domain #3.

XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;

XX major histocompatibility complex; MHC; therapy; immune response;

XX malignancy; chicken.

XX Gallus gallus.

XX Key Location/Qualifiers

XX 58..65 /note= "MHC class I epitope"

XX 66..81 /note= "MHC class II epitope"

XX 93..100 /note= "Javelin sequence"

XX WO200179259-A1.

XX 25-OCT-2001.

XX 17-APR-2001; 2001WO-US12567.

XX 17-APR-2000; 2000US-197462P.

XX (ROTH/) ROTHMAN J E.

XX (MAYH/) MAYHEW M.

XX (HOEW/) HOE M.

XX Rothman JE, Mayhew M, Hoe M;

XX WPI; 2002-017594/02.

XX A new antigenic complex comprising epitopes non-covalently joined to a

XX heat shock protein by a molecular tether designated a javelin are

XX useful to treat or prevent infectious disease or malignancy -

XX Disclosure; Page 14; 47pp; English.

XX The present invention relates to an antigenic complex, comprising a

XX number of epitopes non-covalently joined to a heat shock protein (HSP) by

XX a tethering molecule referred to as javelin which has affinity for the

XX HSP under physiological conditions, where the epitopes are covalently

XX joined to the tethering molecule and one epitope is major

XX histocompatibility complex class I (MHC) and the other MHC class II. The

XX antigenic complex is used to induce immune responses directed towards the

XX treatment or prevention of infectious diseases and malignancies. The

XX present sequence is chicken ovalbumin derived protein domain.

SQ Sequence 100 AA;

Query Match 100.0%; Score 38; DB 23; Length 100;

Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8

|||||

DB 58 SIINFEKL 65

RESULT 113

AAE13459

ID AAE13459 standard; Protein; 103 AA.

PT joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
 PT response or as a treatment or prophylactic against diseases -
 XX
 XX Disclosure; Page 17; 65pp; English.

CC The specification describes a peptide variant of lemA, comprising a
 CC hydrophobic element joined to a CD8+ epitope. The peptides may be
 CC used therapeutically by administering the peptides to a patient having
 CC a need to induce a directed CD8+ T cell response. The peptide may also
 CC be used as a preventive measure to avoid a disease or condition, or to
 CC treat subjects already afflicted with a disease. The present sequence
 CC represents a peptide of the invention.

XX Sequence 43 AA;

Query Match 100.0%; Score 38; DB 22; Length 43;
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFEKL 8
 |||||
 Db 28 SIINFEKL 35

RESULT 107

AAB84321
 ID AAB84321 standard; peptide; 47 AA.

XX AAB84321;

DT 22-AUG-2001 (first entry)

DE Amino acid sequence of a lems variant peptide.

KW lemA; CD8+ epitope; T cell response.

XX Synthetic.

PN WO200140275-A2.

PD 07-JUN-2001.

PF 06-DEC-2000; 2000WO-US33027.

PR 06-DEC-1999; 99US-0169227.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Kurlander RJ, Chao E, Fields J;

DR WPI; 2001-389952/41.

PT New isolated variant of lemA, tlema, comprising a hydrophobic element
 PT joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
 PT response or as a treatment or prophylactic against diseases -

XX Disclosure; Page 19; 65pp; English.

CC The specification describes a peptide variant of lemA, comprising a
 CC hydrophobic element joined to a CD8+ epitope. The peptides may be
 CC used therapeutically by administering the peptides to a patient having
 CC a need to induce a directed CD8+ T cell response. The peptide may also
 CC be used as a preventive measure to avoid a disease or condition, or to
 CC treat subjects already afflicted with a disease. The present sequence
 CC represents a peptide of the invention.

XX Sequence 47 AA;

Query Match 100.0%; Score 38; DB 22; Length 47;
 Best Local Similarity 100.0%; Pred. No. 0.76;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFEKL 8

Db 34 SIINFEKL 41
 |||||

RESULT 108

AAB84322
 ID AAB84322 standard; peptide; 48 AA.

XX AAB84322;

DT 22-AUG-2001 (first entry)

DE Amino acid sequence of a sema variant peptide.

KW lemA; CD8+ epitope; T cell response.

XX Synthetic.

PN WO200140275-A2.

PD 07-JUN-2001.

PF 06-DEC-2000; 2000WO-US33027.

PR 06-DEC-1999; 99US-0169227.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Kurlander RJ, Chao E, Fields J;

DR WPI; 2001-389952/41.

PT New isolated variant of lemA, tlema, comprising a hydrophobic element
 PT joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
 PT response or as a treatment or prophylactic against diseases -

XX Disclosure; Page 19; 65pp; English.

CC The specification describes a peptide variant of lemA, comprising a
 CC hydrophobic element joined to a CD8+ epitope. The peptides may be
 CC used therapeutically by administering the peptides to a patient having
 CC a need to induce a directed CD8+ T cell response. The peptide may also
 CC be used as a preventive measure to avoid a disease or condition, or to
 CC treat subjects already afflicted with a disease. The present sequence
 CC represents a peptide of the invention.

XX Sequence 48 AA;

Query Match 100.0%; Score 38; DB 22; Length 48;
 Best Local Similarity 100.0%; Pred. No. 0.78;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFEKL 8

Db 1 SIINFEKL 8

RESULT 109

AAB48953
 ID AAB48953 standard; Protein; 49 AA.

XX AAB48953;

DT 27-MAR-2001 (first entry)

DE Tn5-DICE ovalbumin MHC class I epitope fusion protein.

KW Transposable element; MHC epitope; major histocompatibility complex;
 KW intracellular bacterial pathogen; loxp site; Cre recombinase;
 KW insertion end; in-frame fusion; detection; antigen;
 KW disseminated insertions of class-I epitopes; DICE-I; transposon Tn5;
 KW ovalbumin MHC class I epitope.

PR 23-MAY-1995; 95US-0447761.
 XX (SCRI) SCRIPPS RES INST.
 XX DeBruijn MLH, Jackson MR, Peterson PA;
 XX WPI; 1997-020850/02.
 DR
 XX
 XX Prodn. of activated CD8+ T cells directed to specific antigen - can
 PT specifically kill target cells useful to treat, e.g. cancer
 PT
 XX
 XX Example 1; Page 26; 84pp; English.
 PS
 XX The method for the production of activated CD8+ T cells specifically
 CC directed towards a particular antigen involves affixing peptides
 CC corresponding to the particular antigen to an artificial support;
 CC contacting macrophages with the affixed peptides for a time sufficient
 CC for the peptides to be engulfed, and at least a portion of the peptides
 CC to be presented on the surface of the macrophage; and contacting
 CC unprimed CD8+ T cells with the peptide presenting macrophages for a
 CC time sufficient to activate the unprimed CD8+ T cells. The present
 CC sequence represents a peptide designated OVA35 which corresponds to
 CC ovalbumin, a Kb-restricted peptide antigen. This is not as efficient as
 CC the optimal peptide. The method, macrophages and artificial antigen
 CC presenting cell, having a peptide corresponding to the particular
 CC antigen present on its surface and at least a portion of an artificial
 CC support in its interior, can be used to treat conditions (e.g. cancer,
 CC tumours, neoplasia, viral or retroviral infection or autoimmune or
 CC autoimmune-type conditions) in patients via the specific killing of
 CC target cells.
 XX
 XX
 SQ Sequence 35 AA;
 Query Match 100.0%; Score 38; DB 19; Length 35;
 Best Local Similarity 100.0%; Pred. NO. 0.56;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 |||||
 Db 16 SIINFEKL 23
 RESULT 105
 AAO26741
 ID AAO26741 standard; Peptide; 36 AA.
 XX
 AC AAO26741;
 XX
 DT 27-MAR-2003 (first entry)
 XX
 XX Chicken ovalbumin class I and II carrying antigenic peptide.
 DE
 XX Cytostatic; universal polypeptidic carrier; GB3 receptor; cytotoxic drug;
 KW tumour cell; immunogenic; gene therapy; chicken.
 XX
 XX Gallus sp.
 OS
 XX EP1229045-A1.
 PN
 XX 07-AUG-2002.
 PD
 XX
 XX 01-FEB-2001; 2001EP-0400255.
 PF
 XX 01-FEB-2001; 2001EP-0400255.
 PR
 XX (CURI-) INST CURIE.
 PA (CNRS) CENT NAT RECH SCI.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (UYPA-) UNIV CURIE PARIS VI P & M.
 XX
 XX Johannes L, Tartour E, Goud B, Fridman WH;
 PI WPI; 2003-185783/19.
 XX
 XX

XX Novel universal polypeptidic carrier for targeting a molecule directly
 PT or indirectly to GB3 receptor expressing cells, useful for directing
 PT cytotoxic drugs to tumor cells expressing GB3 receptor -
 XX
 XX Example 3; Page 8; 24pp; English.
 PS
 XX The invention relates to a universal polypeptidic carrier for targeting
 CC directly or indirectly a molecule to GB3 receptor expressing cells. The
 CC universal polypeptidic carrier is useful for targeting directly or
 CC indirectly, to GB3 receptor expressing cells, using molecules such as
 CC proteins, peptides, oligopeptides, glycoproteins, glycopeptides, nucleic
 CC acids, polynucleotides, or its combination, an antigen to be targeted
 CC antigen presenting cells or a cytotoxic drug or pro-drug to be targeted
 CC to tumour cells expressing GB3 receptor. The universal polypeptidic
 CC carrier is useful for delivering an expression vector containing a
 CC sequence of interest into a GB3 receptor expressing cells which involves
 CC operably linking the expression vector to a lysine-rich peptide
 CC covalently linked to the Cys moiety of the universal polypeptidic
 CC carrier. The lysine rich peptide is preferably a 16-mex poly-lysine and
 CC the sequence of interest is preferably a sequence encoding an immunogenic
 CC peptide, or a sequence encoding a drug or a pro-drug becoming toxic for
 CC the GB3 receptor expressing cells, or a sequence encoding a therapeutic
 CC active molecule. The universal polypeptidic carrier is useful for
 CC targeting a molecule to a GB3 receptor expressing cell to enable the
 CC molecule to be internalised, processed and/or expressed in the cell
 CC expressing a GB3 receptor. The universal polypeptidic carrier is useful
 CC as a carrier for introducing a nucleotide sequence in a target cell
 CC either for gene therapy or for obtaining recombinant cells expressing
 CC heterologous proteins. This sequence represents a chicken ovalbumin class
 CC I and II carrying antigenic peptide relating to the universal carrier
 CC protein of the invention.
 XX
 SQ Sequence 36 AA;
 Query Match 100.0%; Score 38; DB 24; Length 36;
 Best Local Similarity 100.0%; Pred. NO. 0.58;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 |||||
 Db 6 SIINFEKL 13
 RESULT 106
 AAB84325
 ID AAB84325 standard; peptide; 43 AA.
 XX
 AC AAB84325;
 XX
 DT 22-AUG-2001 (first entry)
 XX
 XX Amino acid sequence of a lems variant peptide.
 DE
 XX lema; CD8+ epitope; T cell response.
 KW
 XX Synthetic.
 OS
 XX WO200140275-A2.
 PN
 XX 07-JUN-2001.
 PD
 XX
 XX 06-DEC-2000; 2000WO-US33027.
 PF
 XX 06-DEC-1999; 99US-0169227.
 PR
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Kurlander RJ, Chao E, Fields J;
 XX WPI; 2001-389952/41.
 DR
 XX New isolated variant of lema, tlema, comprising a hydrophobic element
 PT

```

XX FH Key Location/Qualifiers
XX FT Region 1..8
XX FT /note= "Javelin sequence"
XX FT Region 9..11
XX FT /note= "Linker"
XX FT Region 12..19
XX FT /note= "MHC class I epitope"
XX FT Region 20..22
XX FT /note= "Linker"
XX FT Region 23..30
XX FT /note= "Javelin sequence"
XX WO200179259-A1.
XX PD 25-OCT-2001.
XX PF 17-APR-2001; 2001WO-US12567.
XX PR 17-APR-2000; 2000US-197462P.
XX PA (ROTH/) ROTHMAN J E.
XX PA (MAYH/) MAYHEW M.
XX PA (HOEM/) HOE M.
XX PI Rothman JE, Mayhew M, Hoe M;
XX WPI; 2002-017594/02.
XX A new antigenic complex comprising epitopes non-covalently joined to a
XX heat shock protein by a molecular tether designated a javelin are
XX useful to treat or prevent infectious disease or malignancy -
XX Disclosure; Page 13; 47pp; English.
XX The present invention relates to an antigenic complex, comprising a
XX number of epitopes non-covalently joined to a heat shock protein (HSP) by
XX a tethering molecule referred to as javelin which has affinity for the
XX HSP under physiological conditions, where the epitopes are covalently
XX joined to the tethering molecule and one epitope is major
XX histocompatibility complex class I (MHC) and the other MHC class II. The
XX antigenic complex is used to induce immune responses directed towards the
XX treatment or prevention of infectious diseases and malignancies. The
XX present sequence is chicken MHC class I peptide antigen.
XX Sequence 30 AA;
XX Query Match 100.0%; Score 38; DB 23; Length 30;
XX Best Local Similarity 100.0%; Pred. No. 0.48;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 SIINFEKL 8
XX DB 12 SIINFEKL 19
XX RESULT 103
XX ABP57407
XX AC ABP57407;
XX DT 23-APR-2003 (first entry)
XX DE Synthetic 31mer peptide.
XX Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
XX Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
XX virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX Synthetic.
XX WO2003000899-A1.
XX PN

```

```

XX PD 03-JAN-2003.
XX PF 20-JUN-2002; 2002WO-GB02829.
XX PR 22-JUN-2001; 2001GB-0015382.
XX PA (UYER-) UNIV BRISTOL.
XX PI Hirst TR;
XX WPI; 2003-175291/17.
XX Use of a mutant form of B subunit of Escherichia coli heat labile
XX enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX target cell for treating viral infection or cancer -
XX Example 5; Page 45; 84pp; English.
XX The present invention describes a mutant form of B subunit of Escherichia
XX coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX from Vibrio cholerae which is useful for delivering an agent to a target
XX cell, and has GM-1 ganglioside receptor binding activity but has reduced
XX immunogenic and immunomodulatory activity relative to the wild-type form
XX of EtxB or CtxB. Also described: (1) treating a disease or condition in
XX a subject; (2) delivering the agent using the mutant to a target cell;
XX (3) a composition; and (4) a kit for delivering the agent to a target
XX cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
XX can be used in vaccines. The mutant can be used for the preparation of
XX a medicament for delivering an exogenous peptide, which is the agent,
XX into the major histocompatibility complex (MHC) Class I antigen
XX processing and presenting pathways to elicit a cytotoxic T lymphocyte
XX (CTL) response, or for separate, simultaneous or combined use for
XX treating viral infection or cancer. The mutant form of EtxB or CtxB
XX enters mammalian cells without inducing a potent anti-B-subunit response
XX and immunomodulatory response. It may be linked with an agent to
XX upregulate the presentation of the antigen or antigenic determinant.
XX The present sequence represents a peptide which is used in an example
XX from the present invention.
XX Sequence 31 AA;
XX Query Match 100.0%; Score 38; DB 24; Length 31;
XX Best Local Similarity 100.0%; Pred. No. 0.49;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 SIINFEKL 8
XX DB 19 SIINFEKL 26
XX RESULT 104
XX AAW04646
XX ID AAW04646 standard; peptide; 35 AA.
XX AC AAW04646;
XX DT 01-AUG-1997 (first entry)
XX DE Ovalbumin-derived activated CD8+ T cells epitope OVA35.
XX KW Macrophage; artificial antigen presenting cell; APC; cancer;
XX tumours; neoplasia; viral infection; retroviral infection;
XX autoimmune.
XX OS Synthetic.
XX PN WO9637107-A1.
XX PD 28-NOV-1996.
XX PF 22-MAY-1996; 96WO-US07436.
XX PN

```



```

XX 23-APR-2003 (first entry)
XX Synthetic 26mer peptide.
XX Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
XX Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
XX virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX Synthetic.
XX WO2003000899-A1.
XX 03-JAN-2003.
XX 20-JUN-2002; 2002WO-GB02829.
XX 22-JUN-2001; 2001GB-0015382.
XX (UYBR-) UNIV BRISTOL.
XX Hirst TR;
XX WPI; 2003-175291/17.
XX Use of a mutant form of B subunit of Escherichia coli heat labile
XX enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX target cell for treating viral infection or cancer -
XX Example 5; Page 45; 84pp; English.
XX The present invention describes a mutant form of B subunit of Escherichia
XX coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX from Vibrio cholerae which is useful for delivering an agent to a target
XX cell, and has GM-1 ganglioside receptor binding activity but has reduced
XX immunogenic and immunomodulatory activity relative to the wild-type form
XX of EtxB or CtxB. Also described: (1) treating a disease or condition in
XX a subject; (2) delivering the agent using the mutant to a target cell;
XX (3) a composition; and (4) a kit for delivering the agent to a target
XX cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
XX can be used in vaccines. The mutant can be used for the preparation of
XX a medicament for delivering an exogenous peptide, which is the agent,
XX into the major histocompatibility complex (MHC) Class I antigen
XX processing and presenting pathways to elicit a cytotoxic T lymphocyte
XX treating viral infection or cancer. The mutant form of EtxB or CtxB
XX enters mammalian cells without inducing a potent anti-B-subunit response
XX and immunomodulatory response. It may be linked with an agent to
XX upregulate the presentation of the antigen or antigenic determinant.
XX The present sequence represents a peptide which is used in an example
XX from the present invention.
XX Sequence 26 AA;
XX Query Match 100.0%; Score 38; DB 24; Length 26;
XX Best Local Similarity 100.0%; Pred. No. 0.41;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 SIINFEKL 8
XX |||||
XX 19 SIINFEKL 26

RESULT 101
ABP57406
ID ABP57406 standard; peptide; 26 AA.
XX AC ABP57406;
XX 23-APR-2003 (first entry)
XX Synthetic 26mer* peptide.
XX

```

```

KW Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
KW Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
KW virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX Synthetic.
XX WO2003000899-A1.
XX 03-JAN-2003.
XX 20-JUN-2002; 2002WO-GB02829.
XX 22-JUN-2001; 2001GB-0015382.
XX (UYBR-) UNIV BRISTOL.
XX Hirst TR;
XX WPI; 2003-175291/17.
XX Use of a mutant form of B subunit of Escherichia coli heat labile
XX enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX target cell for treating viral infection or cancer -
XX Example 5; Page 45; 84pp; English.
XX The present invention describes a mutant form of B subunit of Escherichia
XX coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX from Vibrio cholerae which is useful for delivering an agent to a target
XX cell, and has GM-1 ganglioside receptor binding activity but has reduced
XX immunogenic and immunomodulatory activity relative to the wild-type form
XX of EtxB or CtxB. Also described: (1) treating a disease or condition in
XX a subject; (2) delivering the agent using the mutant to a target cell;
XX (3) a composition; and (4) a kit for delivering the agent to a target
XX cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
XX can be used in vaccines. The mutant can be used for the preparation of
XX a medicament for delivering an exogenous peptide, which is the agent,
XX into the major histocompatibility complex (MHC) Class I antigen
XX processing and presenting pathways to elicit a cytotoxic T lymphocyte
XX (CTL) response, or for separate, simultaneous or combined use for
XX treating viral infection or cancer. The mutant form of EtxB or CtxB
XX enters mammalian cells without inducing a potent anti-B-subunit response
XX and immunomodulatory response. It may be linked with an agent to
XX upregulate the presentation of the antigen or antigenic determinant.
XX The present sequence represents a peptide which is used in an example
XX from the present invention.
XX Sequence 26 AA;
XX Query Match 100.0%; Score 38; DB 24; Length 26;
XX Best Local Similarity 100.0%; Pred. No. 0.41;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 SIINFEKL 8
XX |||||
XX 19 SIINFEKL 26

RESULT 102
AAEL13448
ID AAEL13448 standard; peptide; 30 AA.
XX AC AAEL13448;
XX 12-FEB-2002 (first entry)
XX Chicken MHC class I peptide antigen #3.
XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
XX major histocompatibility complex; MHC; therapy; immune response;
XX malignancy; chicken.
XX Gallus gallus.
XX

```

Query Match 100.0%; Score 38; DB 22; Length 24;
 Best Local Similarity 100.0%; Pred. No. 0.38; 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 DB 5 SIINFEKL 12

RESULT 98
 AAB74439
 ID AAB74439 standard; peptide; 24 AA.
 AC AAB74439;
 XX
 DT 29-MAY-2001 (first entry)
 XX
 DE Myelin basic protein amino acids 84-102 peptide.
 XX
 KW Cytotoxic T-lymphocyte; CTL response; immunostimulation; infection;
 KW cancer.
 XX
 OS Unidentified.
 XX
 XX US6197311-B1.
 XX
 PD 06-MAR-2001.
 XX
 PF 17-FEB-1998; 98US-0024220.
 XX
 PR 07-JUN-1995; 95US-0476674.
 PR 25-JUL-1991; 91US-0735069.
 PR 24-JUL-1992; 92US-0919787.
 PR 07-DEC-1994; 94US-0351001.
 XX
 PA (IDEC-) IDEC PHARM CORP.
 XX
 PI Raychaudhuri S, Rastetter WH, Black A;
 XX
 DR WPI; 2001-256350/26.
 XX
 PT Treating papillomavirus-related tumor or malignancy, involves
 PT administering an antigen formulation substantially free of
 PT immunostimulatory peptides, and comprising human papillomavirus antigen
 PT and microfluidized adjuvant -
 XX
 PS Disclosure; Column 10-11; 22pp; English.
 XX
 CC The present invention describes a method of treating a
 CC papillomavirus-related tumour, involving administering an antigen
 CC formulation capable of inducing a cytotoxic T-lymphocyte (CTL) response
 CC specific to the papillomavirus antigen in the individual. This is useful
 CC in the treatment of cancer and infectious, such as those due to HIV,
 CC bacteria, parasites, influenza, herpes virus and hepatitis viruses. The
 CC present sequence is a peptide used to demonstrate the method of the
 CC invention.

QY 1 SIINFEKL 8
 DB 5 SIINFEKL 12

RESULT 99
 ABG31664
 ID ABG31664 standard; Peptide; 24 AA.
 XX
 AC ABG31664;

XX
 DT 05-NOV-2002 (first entry)
 XX
 DE Ovalbumin (OVA) peptide fragment.
 XX
 KW Ovalbumin; microfluidised antigen; detergent; micelle-forming agent;
 KW biodegradable oil; biocompatible oil; cytotoxic T lymphocyte; HIV;
 KW human immunodeficiency virus; herpes virus; malaria; influenza; cancer;
 KW hepatitis; respiratory syncytial virus; domesticated animal; OVA;
 KW agricultural animal.
 XX
 OS Unidentified.
 XX
 XX US2002039582-A1.
 XX
 PD 04-APR-2002.
 XX
 PF 20-DEC-2000; 2000US-0740003.
 XX
 PR 17-FEB-1998; 98US-0024220.
 PR 25-JUL-1991; 91US-0735069.
 PR 07-DEC-1994; 94US-0351001.
 PR 07-JUN-1995; 95US-0476674.
 PR 29-AUG-1997; 97US-0919787.
 XX
 PA (IDEC-) IDEC PHARM CORP.
 XX
 PI Raychaudhuri S, Rastetter WH, Black A;
 XX
 DR WPI; 2002-607062/65.
 XX
 PT Composition useful for inducing cytotoxic T lymphocyte response in
 PT domesticated animals and humans comprises antigen mixed with
 PT microfluidised antigen formulation which is substantially free of
 PT immunostimulating peptides -
 XX
 PS Disclosure; Page 6; 3lpp; English.
 XX
 CC The invention relates to a composition comprising an antigen mixed with
 CC microfluidised antigen formulation comprising a stabilised detergent, a
 CC micelle-forming agent and a biodegradable/biocompatible oil. The
 CC composition is formulated as a stable oil-in-water emulsion substantially
 CC free of or lacking immunostimulating peptides, and is capable of inducing
 CC specific cytotoxic T lymphocyte response against antigens in vivo. The
 CC composition is useful for treating patients infected with human
 CC immunodeficiency virus (HIV) or herpes virus, and patients suffering from
 CC malaria, influenza, hepatitis, cancer or respiratory syncytial virus by
 CC administering a composition comprising HIV antigen, malaria-associated
 CC antigen, hepatitis-associated antigen, cancer-associated antigen, herpes
 CC antigen or respiratory syncytial antigen respectively, mixed with
 CC microfluidised antigen consisting essentially of two of stabilising
 CC detergent; micelle-forming agent and biodegradable and biocompatible oil,
 CC the antigen formulation being formulated as stable oil-in-water emulsion,
 CC and inducing cytotoxic T lymphocyte response in a patient e.g. human,
 CC domesticated animal or agricultural animal. This sequence represents an
 CC ovalbumin (OVA) peptide fragment used in the scope of the invention.

QY 1 SIINFEKL 8
 DB 5 SIINFEKL 12

RESULT 100
 ABP57405
 ID ABP57405 standard; peptide; 26 AA.
 XX
 AC ABP57405;

XX 19-FEB-1992; 92US-0841662.
 XX (SCRI) SCRIPPS RES INST.
 XX Jackson M, Langlade-demoyen P, Peterson PA;
 XX WPI; 1993-288401/36.
 XX Prodn. and use of human class I MHC molecules for activation of
 PT CD8 cells - for therapy of e.g. cancer, viral, retroviral and
 PT auto-immune diseases
 XX Disclosure; Page 77; 182pp; English.
 XX Human class I MHC genes are inserted into a cell and placed under
 CC the control of an inducible promoter. This provides a means of
 CC producing, loading and using Class I MHC molecules to specifically
 CC activate CD8 cells in vitro. Activated cells can be used to
 CC specifically kill target cells and also to treat cancer as well as
 CC viral, retroviral, autoimmune and autoimmune-type diseases. When
 CC conjugated to a toxin, empty human MHC molecules expressed by the
 CC cells can be used to inhibit transplant rejection. A number of
 CC antigenic peptides (AAR41450-R41463) are synthesised to be bound by
 CC the MHC molecules and this binding can then activate the CD8 cells.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX Sequence 24 AA;
 SQ Query Match 100.0%; Score 38; DB 14; Length 24;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB |||||||
 5 SIINFEKL 12

RESULT 96
 AAW04645
 ID AAW04645 standard; peptide; 24 AA.
 XX AAW04645;
 XX 01-AUG-1997 (first entry)
 XX Ovalbumin-derived activated CD8+ T cells epitope OVA24.
 XX Macrophage; artificial antigen presenting cell; APC; cancer;
 XX tumours; neoplasia; viral infection; retroviral infection;
 XX autoimmune.
 XX Synthetic.
 XX WO9637107-A1.
 XX 28-NOV-1996.
 XX 22-MAY-1996; 96WO-US07436.
 XX 23-MAY-1995; 95US-0447761.
 XX (SCRI) SCRIPPS RES INST.
 XX DeBruijn MLH, Jackson MR, Peterson PA;
 XX WPI; 1997-020850/02.
 XX Prodn. of activated CD8+ T cells directed to specific antigen - can
 PT specifically kill target cells useful to treat, e.g. cancer
 XX Example 1; Page 26; 84pp; English.
 XX

CC The method for the production of activated CD8+ T cells specifically
 CC directed towards a particular antigen involves affixing peptides
 CC corresponding to the particular antigen to an artificial support;
 CC contacting macrophages with the affixed peptides for a time sufficient
 CC for the peptides to be engulfed, and at least a portion of the peptides
 CC to be presented on the surface of the macrophage; and contacting
 CC unprimed CD8+ T cells with the peptide presenting macrophages for a
 CC time sufficient to activate the unprimed CD8+ T cells. The present
 CC sequence represents a peptide designated OVA24 which corresponds to
 CC ovalbumin, a Kb-restricted peptide antigen. This is not as efficient as
 CC the optimal peptide. The method, macrophages and artificial antigen
 CC presenting cell, having a peptide corresponding to the particular
 CC antigen present on its surface and at least a portion of an artificial
 CC support in its interior, can be used to treat conditions (e.g. cancer,
 CC tumours, neoplasia, viral or retroviral infection or autoimmune or
 CC autoimmune-type conditions) in patients via the specific killing of
 CC target cells.
 XX Sequence 24 AA;
 SQ Query Match 100.0%; Score 38; DB 18; Length 24;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB |||||||
 5 SIINFEKL 12

RESULT 97
 AAG65170
 ID AAG65170 standard; peptide; 24 AA.
 XX AAG65170;
 XX 15-OCT-2002 (first entry)
 XX Ovalbumin based peptide.
 XX Cytotoxic T-lymphocyte response; CTL; antigen; viral infection;
 XX bacterial infection; cancer; parasitic infection; immune response;
 XX non-toxic.
 XX Unidentified.
 XX US6270769-B1.
 XX 07-AUG-2001.
 XX 24-MAY-1995; 95US-0449728.
 XX 24-JUL-1992; 92US-0919787.
 XX 25-JUL-1991; 91US-0735069.
 XX (IDEC-) IDEC PHARM CORP.
 XX Raychaudhuri S, Rastetter WH;
 XX WPI; 2001-564234/63.
 XX Induction of cytotoxic T-lymphocyte responses -
 XX Disclosure; Column 17-18; 24pp; English.
 XX The present invention relates to a method of treating viral, parasitic
 CC and bacterial infections and cancer in humans, by administering an
 CC antigen which causes a cytotoxic T-lymphocyte response. Said antigen does
 CC not contain an immunostimulatory element. The method can also be used in
 CC domesticated animals. The present sequence is a peptide used as an
 CC antigen in the exemplification of the invention.
 XX Sequence 24 AA;
 SQ

XX ABP57404;
 AC
 XX
 DT 23-APR-2003 (first entry)
 DE
 XX
 XX Synthetic 19mer peptide.
 XX
 KW Escherichia coli heat labile enterotoxin; EtX; cholera toxin; Ctx;
 KW Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
 KW virucide; cytostatic; vaccine; viral infection; cancer; EtXB; CtxB.
 XX
 OS Synthetic.
 XX
 XX WO2003000899-A1.
 PN
 XX
 PD 03-JAN-2003.
 XX
 XX 20-JUN-2002; 2002WO-GB02829.
 PF
 XX
 PR 22-JUN-2001; 2001GB-0015382.
 XX
 PA (UYER-) UNIV BRISTOL.
 XX
 XX Hirst TR;
 PI
 XX
 XX WPI; 2003-175291/17.
 DR
 XX
 PT Use of a mutant form of B subunit of Escherichia coli heat labile
 PT enterotoxin or B subunit of cholera toxin for delivering an agent to a
 PT target cell for treating viral infection or cancer -
 XX
 PS Example 5; Page 45; 84pp; English.
 XX
 CC The present invention describes a mutant form of B subunit of Escherichia
 CC coli heat labile enterotoxin (EtXB) or B subunit of cholera toxin (CtxB)
 CC from Vibrio cholerae which is useful for delivering an agent to a target
 CC cell, and has GM-1 ganglioside receptor binding activity but has reduced
 CC immunogenic and immunomodulatory activity relative to the wild-type form
 CC of EtXB or CtxB. Also described: (1) treating a disease or condition in
 CC a subject; (2) delivering the agent using the mutant to a target cell;
 CC (3) a composition; and (4) a kit for delivering the agent to a target
 CC cell. Mutant EtXB and CtxB have virucide and cytostatic activities and
 CC can be used in vaccines. The mutant can be used for the preparation of
 CC a medicament for delivering an exogenous peptide, which is the agent,
 CC into the major histocompatibility complex (MHC) Class I antigen
 CC processing and presenting pathways to elicit a cytotoxic T lymphocyte
 CC (CTL) response, or for separate, simultaneous or combined use for
 CC treating viral infection or cancer. The mutant form of EtXB or CtxB
 CC enters mammalian cells without inducing a potent anti-B-subunit response
 CC and immunomodulatory response. It may be linked with an agent to
 CC upregulate the presentation of the antigen or antigenic determinant.
 CC The present sequence represents a peptide which is used in an example
 CC from the present invention.
 XX
 SQ Sequence 19 AA;
 Query Match 100.0%; Score 38; DB 24; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB |||||
 12 SIINFEKL 19
 RESULT 94
 AAR32294
 ID AAR32294 standard; protein; 24 AA.
 XX
 AC AAR32294;
 XX
 XX 25-MAR-2003 (updated)
 DT 31-MAY-1993 (first entry)
 DT

XX Sequence of synthetic peptide ova 253-276 which corresp. to AAs 253-
 DE 276 of ovalbumin.
 DE
 XX Cytotoxic T lymphocyte response; epitope; antigen.
 KW
 XX Synthetic.
 OS
 XX WO9301831-A1.
 PN
 XX 04-FEB-1993.
 PD
 XX 24-JUL-1992; 92WO-US06193.
 PF
 XX 25-JUL-1991; 91US-0735069.
 PR
 XX (IDEC-) IDEC PHARM CORP.
 PA
 XX Rastetter WH, Raychaudhuri S;
 PI
 XX WPI; 1993-058526/07.
 DR
 XX New compsn. comprising an antigen and a formulation - to induce a
 PT cytotoxic T-lymphocyte response, useful for treating malaria,
 PT HIV, influenza, hepatitis, herpes, cancer, etc.
 XX
 PS Disclosure; Page 19; 56pp; English.
 XX
 CC Carbone and Bevan demonstrated that cytotoxic T-lymphocyte (CTL)
 CC induced in C57BL/6 mice by EG7-ova transfectant, and by
 CC cytoplasmically ova-loaded splenocytes recognise EL4 cells coated
 CC with the peptide ova 258-276. To determine whether soluble ovalbumin
 CC in AF induces similar CTL responses, spleen cells were prepared from
 CC immunised mice and stimulated in vitro with EG7-ova. The effectors
 CC were tested against EL4 cells coated with the peptide ova 253-276
 CC or with a control peptide derived from myelin basic protein (MBP 84-
 CC 102). The results demonstrate that ova-AF primed CTL with a similar
 CC specificity to those primed by transfectants, or by cytoplasmically
 CC loaded ova.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 24 AA;
 Query Match 100.0%; Score 38; DB 14; Length 24;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB |||||
 5 SIINFEKL 12
 RESULT 95
 AAR41450
 ID AAR41450 standard; Protein; 24 AA.
 XX
 AC AAR41450;
 XX
 DT 25-MAR-2003 (updated)
 DT 23-FEB-1994 (first entry)
 XX
 DE Antigenic peptide bound by MHC class one molecules.
 XX
 KW HLA; Human Leucocyte Antigen; MHC; Class one molecules; cancer;
 KW autoimmunity; transplant rejection; T-cell activation.
 XX
 OS Synthetic.
 XX
 PN WO9317095-A1.
 XX
 XX 02-SEP-1993.
 PD
 XX 18-FEB-1993; 93WO-US01557.
 PF

Query Match 100.0%; Score 38; DB 18; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFEKL 8
 |||||
 Db 1 SIINFEKL 8

RESULT 91

AAE13446
 ID AAE13446 standard; peptide; 19 AA.

XX AC AAE13446;

DT 12-FEB-2002 (first entry)

XX Chicken MHC class I peptide antigen #1.

XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
 KW major histocompatibility complex; MHC; therapy; immune response;
 KW malignancy; chicken.

XX OS Gallus gallus.

FT Key Location/Qualifiers
 FT Region 1..8
 FT /note= "MHC class I epitope"
 FT Region 9..11
 FT /note= "Linker"
 FT Region 12..19
 FT /note= "Javelin sequence"

XX WO200179259-A1.

XX 25-OCT-2001.

XX 17-APR-2001; 2001WO-US12567.

XX 17-APR-2000; 2000US-197462P.

XX (ROTH/) ROTHMAN J E.

XX (MAYH/) MAYHEW M.

XX (HOEM/) HOE M.

XX Rothman JE, Mayhew M, Hoe M;

XX WPI; 2002-017594/02.

XX A new antigenic complex comprising epitopes non-covalently joined to a
 PT heat shock protein by a molecular tether designated a javelin are
 PT useful to treat or prevent infectious disease or malignancy -

XX Disclosure; Page 13; 47pp; English.

XX The present invention relates to an antigenic complex, comprising a
 CC number of epitopes non-covalently joined to a heat shock protein (HSP) by
 CC a tethering molecule referred to as javelin which has affinity for the
 CC HSP under physiological conditions, where the epitopes are covalently
 CC joined to the tethering molecule and one epitope is major
 CC histocompatibility complex class I (MHC) and the other MHC class II. The
 CC antigenic complex is used to induce immune responses directed towards the
 CC treatment or prevention of infectious diseases and malignancies. The
 CC present sequence is chicken MHC class I peptide antigen.

XX Sequence 19 AA;

Query Match 100.0%; Score 38; DB 23; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFEKL 8
 |||||

Db 1 SIINFEKL 8

RESULT 92

AAE13447

ID AAE13447 standard; peptide; 19 AA.

XX AC AAE13447;

DT 12-FEB-2002 (first entry)

XX Chicken MHC class I peptide antigen #2.

XX Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
 KW major histocompatibility complex; MHC; therapy; immune response;
 KW malignancy; chicken.

XX OS Gallus gallus.

FT Key Location/Qualifiers
 FT Region 1..8
 FT /note= "Javelin sequence"
 FT Region 9..11
 FT /note= "Linker"
 FT Region 12..19
 FT /note= "MHC class I epitope"

XX WO200179259-A1.

XX 25-OCT-2001.

XX 17-APR-2001; 2001WO-US12567.

XX 17-APR-2000; 2000US-197462P.

XX (ROTH/) ROTHMAN J E.

XX (MAYH/) MAYHEW M.

XX (HOEM/) HOE M.

XX Rothman JE, Mayhew M, Hoe M;

XX WPI; 2002-017594/02.

XX A new antigenic complex comprising epitopes non-covalently joined to a
 PT heat shock protein by a molecular tether designated a javelin are
 PT useful to treat or prevent infectious disease or malignancy -

XX Disclosure; Page 13; 47pp; English.

XX The present invention relates to an antigenic complex, comprising a
 CC number of epitopes non-covalently joined to a heat shock protein (HSP) by
 CC a tethering molecule referred to as javelin which has affinity for the
 CC HSP under physiological conditions, where the epitopes are covalently
 CC joined to the tethering molecule and one epitope is major
 CC histocompatibility complex class I (MHC) and the other MHC class II. The
 CC antigenic complex is used to induce immune responses directed towards the
 CC treatment or prevention of infectious diseases and malignancies. The
 CC present sequence is chicken MHC class I peptide antigen.

XX Sequence 19 AA;

Query Match 100.0%; Score 38; DB 23; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SIINFEKL 8
 |||||

Db 12 SIINFEKL 19

RESULT 93

ABP57404

ID ABP57404 standard; peptide; 19 AA.

CC cell, and has GM-1 ganglioside receptor binding activity but has reduced
 CC immunogenic and immunomodulatory activity relative to the wild-type form
 CC of BtxB or CtxB. Also described: (1) treating a disease or condition in
 CC a subject; (2) delivering the agent using the mutant to a target cell;
 CC (3) a composition; and (4) a kit for delivering the agent to a target
 CC cell. Mutant BtxB and CtxB have virucide and cytostatic activities and
 CC can be used in vaccines. The mutant can be used for the preparation of
 CC a medicament for delivering an exogenous peptide, which is the agent,
 CC into the major histocompatibility complex (MHC) Class I antigen
 CC processing and presenting pathways to elicit a cytotoxic T lymphocyte
 CC (CTL) response, or for separate, simultaneous or combined use for
 CC treating viral infection or cancer. The mutant form of BtxB or CtxB
 CC enters mammalian cells without inducing a potent anti-B-subunit response
 CC and immunomodulatory response. It may be linked with an agent to
 CC upregulate the presentation of the antigen or antigenic determinant.
 CC The present sequence represents a peptide which is used in an example
 CC from the present invention.

XX
 XX SQ Sequence 16 AA;

Query Match 100.0%; Score 38; DB 24; Length 16;
 Best Local Similarity 100.0%; Pred. No. 0.25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 Db 9 SIINFEKL 16

RESULT 89
 AAW19957
 ID AAW19957 standard; Peptide; 19 AA.
 XX
 XX AC AAW19957;
 XX
 XX DT 10-NOV-1997 (first entry)
 XX
 XX DE BiP-binding domain-OVA hybrid peptide.
 XX
 XX KW Vaccine; immunotherapy; heat shock protein; BiP; OVA; cancer;
 XX infectious disease.
 XX OS Synthetic.

FH Key Location/Qualifiers
 FT Peptide 1..8
 FT /label= BiP
 FT Peptide 9..11
 FT /label= Linker
 FT Peptide 12..19
 FT /label= Ova

PN WO9706821-A1.
 XX
 XX DT 27-FEB-1997.
 XX
 XX PF 16-AUG-1996; 96WO-US13363.
 XX
 XX PR 18-AUG-1995; 95US-0002490.
 XX 18-AUG-1995; 95US-0002479.
 XX
 XX PA (SLOK) SLOAN KETTERING INST CANCER RES.
 XX
 XX PI Hartl FU, Hoe MH, Houghton A, Mayhew M, Rothman JE;
 XX Takeuchi Y;
 XX WPI; 1997-165035/15.
 XX
 XX Compn. for inducing immune response contg. antigen and heat shock
 XX protein - also new hybrid peptide and related nucleic acid, for
 XX treatment of infectious diseases and tumours

PS Example 1; Page 18; 58pp; English.

XX Hybrid peptides OVA-BiP (AAW19956) and BiP-OVA (AAW19957) comprise
 CC chicken OVA-peptide (see AAW19955) joined via a peptide linker to
 CC heat shock protein (HSP) BiP binding domain (see also AAW19951).
 CC The hybrid peptide can be combined in vitro with a HSP, such as
 CC hsp70, to form a complex that, when administered to a subject,
 CC induces an immune response.

XX SQ Sequence 19 AA;

Query Match 100.0%; Score 38; DB 18; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 Db 12 SIINFEKL 19

RESULT 90
 AAW19956
 ID AAW19956 standard; Peptide; 19 AA.
 XX
 XX AC AAW19956;
 XX
 XX DT 10-NOV-1997 (first entry)
 XX
 XX DE OVA-BiP-binding domain hybrid peptide.
 XX
 XX KW Vaccine; immunotherapy; heat shock protein; BiP; OVA; cancer;
 XX infectious disease.
 XX OS Synthetic.

FH Key Location/Qualifiers
 FT Peptide 1..8
 FT /label= OVA
 FT Peptide 9..11
 FT /label= Linker
 FT Peptide 12..19
 FT /label= BiP

PN WO9706821-A1.
 XX
 XX DT 27-FEB-1997.
 XX
 XX PF 16-AUG-1996; 96WO-US13363.
 XX
 XX PR 18-AUG-1995; 95US-0002490.
 XX 18-AUG-1995; 95US-0002479.
 XX
 XX PA (SLOK) SLOAN KETTERING INST CANCER RES.
 XX
 XX PI Hartl FU, Hoe MH, Houghton A, Mayhew M, Rothman JE;
 XX Takeuchi Y;
 XX WPI; 1997-165035/15.
 XX
 XX Compn. for inducing immune response contg. antigen and heat shock
 XX protein - also new hybrid peptide and related nucleic acid, for
 XX treatment of infectious diseases and tumours

PS Example 1; Page 18; 58pp; English.

XX Hybrid peptides OVA-BiP (AAW19956) and BiP-OVA (AAW19957) comprise
 CC chicken OVA-peptide (see AAW19955) joined via a peptide linker to
 CC heat shock protein (HSP) BiP binding domain (see also AAW19951).
 CC The hybrid peptide can be combined in vitro with a HSP, such as
 CC hsp70, to form a complex that, when administered to a subject,
 CC induces an immune response.

XX SQ Sequence 19 AA;

```

XX PR 14-APR-2000; 2000AT-0000657.
XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX PI Mattner F, Zauner W, Schmidt W, Buschle M;
XX PS WPI; 2002-025970/03.
XX DR
XX PT Pharmaceutical preparation for use as a potent vaccine for inducing an
XX PT improved immune response in a mammal, comprises a modified peptide -
XX PS Example 4; Page 11; 18pp; English.
XX CC The invention relates to a pharmaceutical preparation comprising a
XX CC modified peptide, which induces an improved immune response in a mammal
XX CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
XX CC negatively charged (Glu), one positively charged (Lys) amino acid) was
XX CC rendered negative by adding (at the N-terminus) Glu or Glu Asp Glu
XX CC Asp, respectively. Results showed that the addition of 4 negatively-
XX CC charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
XX CC this peptide (in combination with poly-L-arginine) able to induce a high
XX CC amount of specific interferon (IFN)-gamma-producing T cells in the
XX CC draining (popliteal) lymph node (local response) and in the spleen
XX CC (systemic response). Thus, the addition of hydrophobic amino acids as
XX CC well as the addition of negatively charged amino acids transforms the
XX CC peptide SIINFEKL to a good inducer of specific T cells. The modified
XX CC peptides of the pharmaceutical composition induce a stronger immune
XX CC response in a mammal compared to wild type antigens. The present
XX CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
XX CC peptide #8 as described in the method of the invention.
XX SQ Sequence 14 AA;
XX
XX Query Match 100.0%; Score 38; DB 23; Length 14;
XX Best Local Similarity 100.0%; Pred. No. 0.22;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 SIINFEKL 8
DB |||||
7 SIINFEKL 14

RESULT 87
AAU09824
ID AAU09824 standard; peptide; 15 AA.
XX AC AAU09824;
XX DT 14-FEB-2002 (first entry)
XX DE Modified ovalbumin-derived class I H-2Kb restricted peptide #4.
XX KW Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
XX KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
XX KW popliteal lymph node; spleen; immune response; systemic response.
XX OS Synthetic.
XX PN WO200178767-A2.
XX PD 25-OCT-2001.
XX PF 17-APR-2001; 2001WO-EP04313.
XX PR 14-APR-2000; 2000AT-0000657.
XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX PI Mattner F, Zauner W, Schmidt W, Buschle M;
XX DR WPI; 2002-025970/03.
XX CC The invention relates to a pharmaceutical preparation comprising a
XX CC modified peptide, which induces an improved immune response in a mammal
XX CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
XX CC negatively charged (Glu), one positively charged (Lys) amino acid) was
XX CC rendered negative by adding (at the N-terminus) Glu or Glu Asp Glu
XX CC Asp, respectively. Results showed that the addition of 4 negatively-
XX CC charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
XX CC this peptide (in combination with poly-L-arginine) able to induce a high
XX CC amount of specific interferon (IFN)-gamma-producing T cells in the
XX CC draining (popliteal) lymph node (local response) and in the spleen
XX CC (systemic response). Thus, the addition of hydrophobic amino acids as
XX CC well as the addition of negatively charged amino acids transforms the
XX CC peptide SIINFEKL to a good inducer of specific T cells. The modified
XX CC peptides of the pharmaceutical composition induce a stronger immune
XX CC response in a mammal compared to wild type antigens. The present
XX CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
XX CC peptide #8 as described in the method of the invention.
XX SQ Sequence 14 AA;
XX
XX Query Match 100.0%; Score 38; DB 23; Length 14;
XX Best Local Similarity 100.0%; Pred. No. 0.22;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 SIINFEKL 8
DB |||||
7 SIINFEKL 14

RESULT 88
ABP57403
ID ABP57403 standard; peptide; 16 AA.
XX AC ABP57403;
XX DT 23-APR-2003 (first entry)
XX DE Synthetic 16mer peptide.
XX KW Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
XX KW Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
XX KW virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX OS Synthetic.
XX PN WO2003000899-A1.
XX PD 03-JAN-2003.
XX PF 20-JUN-2002; 2002WO-GB02829.
XX PR 22-JUN-2001; 2001GB-0015382.
XX PA (UYER-) UNIV BRISTOL.
XX PI Hirst TR;
XX DR WPI; 2003-175291/17.
XX PT Use of a mutant form of B subunit of Escherichia coli heat labile
XX PT enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX PT target cell for treating viral infection or cancer -
XX PS Example 5; Page 45; 84pp; English.
XX CC The present invention describes a mutant form of B subunit of Escherichia
XX CC coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX CC from Vibrio cholerae which is useful for delivering an agent to a target

```

```

PT Pharmaceutical preparation for use as a potent vaccine for inducing an
PT improved immune response in a mammal, comprises a modified peptide -
XX Example 1; Page 9; 18pp; English.
XX
XX The invention relates to a pharmaceutical preparation comprising a
XX modified peptide, which induces an improved immune response in a mammal
XX compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
XX negatively charged (Glu), one positively charged (Lys) amino acid) was
XX rendered negative by adding (at the N-terminus) Glu or Glu Asp Glu
XX Asp, respectively. Results showed that the addition of 4 negatively-
XX charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
XX this peptide (in combination with poly-L-arginine) able to induce a high
XX amount of specific interferon (IFN)-gamma-producing T cells in the
XX draining (popliteal) lymph node (local response) and in the spleen
XX (systemic response). Thus, the addition of hydrophobic amino acids as
XX well as the addition of negatively charged amino acids transforms the
XX peptide SIINFEKL to a good inducer of specific T cells. The modified
XX peptides of the pharmaceutical composition induce a stronger immune
XX response in a mammal compared to wild type antigens. The present
XX sequence represents modified ovalbumin-derived class I H-2Kb restricted
XX peptide #4 as described in the method of the invention.
XX SQ Sequence 15 AA;
XX
XX Query Match 100.0%; Score 38; DB 23; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 0.23;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 SIINFEKL 8
DB |||||
8 SIINFEKL 15

RESULT 89
ABP57403
ID ABP57403 standard; peptide; 16 AA.
XX AC ABP57403;
XX DT 23-APR-2003 (first entry)
XX DE Synthetic 16mer peptide.
XX KW Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
XX KW Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
XX KW virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX OS Synthetic.
XX PN WO2003000899-A1.
XX PD 03-JAN-2003.
XX PF 20-JUN-2002; 2002WO-GB02829.
XX PR 22-JUN-2001; 2001GB-0015382.
XX PA (UYER-) UNIV BRISTOL.
XX PI Hirst TR;
XX DR WPI; 2003-175291/17.
XX PT Use of a mutant form of B subunit of Escherichia coli heat labile
XX PT enterotoxin or B subunit of cholera toxin for delivering an agent to a
XX PT target cell for treating viral infection or cancer -
XX PS Example 5; Page 45; 84pp; English.
XX CC The present invention describes a mutant form of B subunit of Escherichia
XX CC coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
XX CC from Vibrio cholerae which is useful for delivering an agent to a target

```

DE Peptide insert in CACTES-Cys-Ova.
 XX
 KW Bordetella pertussis; adenylate cyclase; CyaA; adenylcyclase;
 KW vector; drug delivery; antigen delivery; cell targeting; CD11b;
 KW CACTES-Cys-Ova.
 XX
 XX Synthetic.
 OS
 XX EP1188446-A1.
 PN
 XX 20-MAR-2002.
 PD
 XX 15-SEP-2000; 2000EP-0402562.
 PF
 XX 15-SEP-2000; 2000EP-0402562.
 PR
 XX (INSP) INST PASTEUR.
 PA (CNRS) CENT NAT RECH SCI.
 XX
 XX Lelerc C, Guernonprez P, Ladant D, Guiso N, Khelef N;
 PI
 XX WPI; 2002-354020/39.
 DR
 XX
 XX Use of Bordetella adenylcyclase to make proteinaceous vector, useful
 PT for drug or antigen delivery, selectively targets cells that express
 PT CD11b -
 PT
 XX Example A; Page 10; 34pp; English.
 PS
 XX The present sequence is a peptide that was introduced into the
 CC catalytic domain of a detoxified form of the adenylate cyclase
 CC (CyaA, or adenylcyclase) of Bordetella pertussis. A recombinant
 CC detoxified CyaA toxin, CACTES-Cys-Ova, harbouring a unique cysteine
 CC that was derived from the peptide insert, was produced. The
 CC protein was labeled on its unique cysteine, and used to detect
 CC CyaA binding to neutrophils. Experiments showed that CyaA binding
 CC to the surface of 3 myeloid cell lines of mouse or human origin, as
 CC well as to human neutrophils, was mainly mediated through the
 CC CD11b/CD18 integrin. The invention relates to the novel use of
 CC Bordetella CyaA as a proteinaceous vector for targeting a
 CC molecule of interest to the surface CD11b-expressing cells,
 CC especially dendritic cells and neutrophils. The molecule of
 CC interest is translocated in the cytosol to prime a cytotoxic T
 CC lymphocyte response. In a preferred embodiment, a peptide is
 CC inserted into in the catalytic domain of CyaA at a permissive site.
 CC The peptide may be an intracellular bacterial cell, tumour, viral,
 CC fungal or parasite cell antigen (all claimed). Alternatively, a
 CC drug, especially an antiinflammatory, is chemically coupled to
 CC CyaA for drug delivery.
 XX
 XX Sequence 14 AA;
 SQ
 Query Match 100.0%; Score 38; DB 23; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB |||||
 5 SIINFEKL 12
 RESULT 85
 AAU09823
 ID AAU09823 standard; peptide; 14 AA.
 XX
 AC AAU09823;
 XX
 XX 14-FEB-2002 (first entry)
 DT
 XX Modified ovalbumin-derived class I H-2Kb restricted peptide #3.
 DE
 XX Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
 KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
 KW popliteal lymph node; spleen; immune response; systemic response.
 OS Synthetic.
 XX WO200178767-A2.
 PN
 XX 25-OCT-2001.
 PD
 XX 17-APR-2001; 2001WO-EP04313.
 PF

KW popliteal lymph node; spleen; immune response; systemic response.
 XX Synthetic.
 OS
 XX WO200178767-A2.
 PN
 XX 25-OCT-2001.
 PD
 XX 17-APR-2001; 2001WO-EP04313.
 PF
 XX 14-APR-2000; 2000AT-0000657.
 PR
 XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 PA
 XX Mattner F, Zauner W, Schmidt W, Buschle M;
 PI
 XX WPI; 2002-025970/03.
 DR
 XX Pharmaceutical preparation for use as a potent vaccine for inducing an
 PT improved immune response in a mammal, comprises a modified peptide -
 PT
 XX Example 1; Page 9; 18pp; English.
 PS
 XX The invention relates to a pharmaceutical preparation comprising a
 CC modified peptide, which induces an improved immune response in a mammal
 CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
 CC negatively charged (Glu), one positively charged (Lys) amino acid) was
 CC rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
 CC Asp, respectively. Results showed that the addition of 4 negatively-
 CC charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
 CC this peptide (in combination with poly-L-arginine) able to induce a high
 CC amount of specific interferon (IFN)-gamma-producing T cells in the
 CC draining (popliteal) lymph node (local response) and in the spleen
 CC (systemic response). Thus, the addition of hydrophobic amino acids as
 CC well as the addition of negatively charged amino acids transforms the
 CC peptide SIINFEKL to a good inducer of specific T cells. The modified
 CC peptides of the pharmaceutical composition induce a stronger immune
 CC response in a mammal compared to wild type antigens. The present
 CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
 CC peptide #3 as described in the method of the invention.
 XX
 XX Sequence 14 AA;
 SQ
 Query Match 100.0%; Score 38; DB 23; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SIINFEKL 8
 DB |||||
 7 SIINFEKL 14
 RESULT 86
 AAU09828
 ID AAU09828 standard; peptide; 14 AA.
 XX
 AC AAU09828;
 XX
 XX 14-FEB-2002 (first entry)
 DT
 XX Modified ovalbumin-derived class I H-2Kb restricted peptide #8.
 DE
 XX Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
 KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
 KW popliteal lymph node; spleen; immune response; systemic response.
 XX
 OS Synthetic.
 XX WO200178767-A2.
 PN
 XX 25-OCT-2001.
 PD
 XX 17-APR-2001; 2001WO-EP04313.
 PF

Query Match 100.0%; Score 38; DB 23; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 DB 5 SIINFEKL 12

RESULT 82
 AAU09826
 ID AAU09826 standard; peptide; 12 AA.
 XX
 AC AAU09826;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Modified ovalbumin-derived class I H-2Kb restricted peptide #6.
 XX
 KW Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
 KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
 KW popliteal lymph node; spleen; immune response; systemic response.
 XX
 OS Synthetic.
 XX
 PN WO200178767-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 17-APR-2001; 2001WO-EP04313.
 XX
 PR 14-APR-2000; 2000AT-0000657.
 XX
 PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 XX
 PI Mattner F, Zauner W, Schmidt W, Buschle M;
 XX
 DR WPI; 2002-025970/03.
 XX
 PT Pharmaceutical preparation for use as a potent vaccine for inducing an
 PT improved immune response in a mammal, comprises a modified peptide -
 XX
 PS Example 2; Page 9; 18pp; English.
 XX
 CC The invention relates to a pharmaceutical preparation comprising a
 CC modified peptide, which induces an improved immune response in a mammal
 CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
 CC negatively charged (Glu), one positively charged (Lys) amino acid) was
 CC rendered negative by adding (at the N-terminus) Glu or Glu Asp Glu
 CC Asp, respectively. Results showed that the addition of 4 negatively-
 CC charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
 CC this peptide (in combination with poly-L-arginine) able to induce a high
 CC amount of specific interferon (IFN)-gamma-producing T cells in the
 CC draining (popliteal) lymph node (local response) and in the spleen
 CC (systemic response). Thus, the addition of hydrophobic amino acids as
 CC well as the addition of negatively charged amino acids transforms the
 CC peptide SIINFEKL to a good inducer of specific T cells. The modified
 CC peptides of the pharmaceutical composition induce a stronger immune
 CC response in a mammal compared to wild type antigens. The present
 CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
 CC peptide #6 as described in the method of the invention.

QY 1 SIINFEKL 8
 DB 5 SIINFEKL 12

Sequence 12 AA;

Query Match 100.0%; Score 38; DB 23; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 DB 5 SIINFEKL 12

RESULT 83
 AAU09827
 ID AAU09827 standard; peptide; 12 AA.
 XX
 AC AAU09827;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Modified ovalbumin-derived class I H-2Kb restricted peptide #7.
 XX
 KW Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
 KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
 KW popliteal lymph node; spleen; immune response; systemic response.
 XX
 OS Synthetic.
 XX
 PN WO200178767-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 17-APR-2001; 2001WO-EP04313.
 XX
 PR 14-APR-2000; 2000AT-0000657.
 XX
 PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
 XX
 PI Mattner F, Zauner W, Schmidt W, Buschle M;
 XX
 DR WPI; 2002-025970/03.
 XX
 PT Pharmaceutical preparation for use as a potent vaccine for inducing an
 PT improved immune response in a mammal, comprises a modified peptide -
 XX
 PS Example 3; Page 10; 18pp; English.
 XX
 CC The invention relates to a pharmaceutical preparation comprising a
 CC modified peptide, which induces an improved immune response in a mammal
 CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
 CC negatively charged (Glu), one positively charged (Lys) amino acid) was
 CC rendered negative by adding (at the N-terminus) Glu or Glu Asp Glu
 CC Asp, respectively. Results showed that the addition of 4 negatively-
 CC charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
 CC this peptide (in combination with poly-L-arginine) able to induce a high
 CC amount of specific interferon (IFN)-gamma-producing T cells in the
 CC draining (popliteal) lymph node (local response) and in the spleen
 CC (systemic response). Thus, the addition of hydrophobic amino acids as
 CC well as the addition of negatively charged amino acids transforms the
 CC peptide SIINFEKL to a good inducer of specific T cells. The modified
 CC peptides of the pharmaceutical composition induce a stronger immune
 CC response in a mammal compared to wild type antigens. The present
 CC sequence represents modified ovalbumin-derived class I H-2Kb restricted
 CC peptide #7 as described in the method of the invention.

QY 1 SIINFEKL 8
 DB 1 SIINFEKL 8

Sequence 12 AA;

Query Match 100.0%; Score 38; DB 23; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
 DB 1 SIINFEKL 8

RESULT 84
 ABB76049
 ID ABB76049 standard; Peptide; 14 AA.
 XX
 AC ABB76049;
 XX
 DT 12-JUL-2002 (first entry)
 XX

PS Example 2; Page 9; 18pp; English.

XX The invention relates to a pharmaceutical preparation comprising a

CC modified peptide, which induces an improved immune response in a mammal

CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one

CC negatively charged (Glu), one positively charged (Lys) amino acid) was

CC rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu

CC Asp, respectively. Results showed that the addition of 4 negatively-

CC charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes

CC this peptide (in combination with poly-L-arginine) able to induce a high

CC amount of specific interferon (IFN)-gamma-producing T cells in the

CC (draining popliteal) lymph node (local response) and in the spleen

CC (systemic response). Thus, the addition of hydrophobic amino acids as

CC well as the addition of negatively charged amino acids transforms the

CC peptide SIINFEKL to a good inducer of specific T cells. The modified

CC peptides of the pharmaceutical composition induce a stronger immune

CC response in a mammal compared to wild type antigens. The present

CC sequence represents modified ovalbumin-derived class I H-2Kb restricted

CC peptide #5 as described in the method of the invention.

XX

SQ Sequence 10 AA;

Query Match 100.0%; Score 38; DB 23; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8

DB 3 SIINFEKL 10

RESULT 80

AAW14122

ID AAW14122 standard; peptide; 12 AA.

AC AAW14122;

XX 20-OCT-1997 (first entry)

DT OVA protein derived MHC class I binding peptide.

DE Major histocompatibility complex; MHC; target; binding; tumour;

XX cancer; neoplasia; LSTRA; EL-4; identification; detection; screening;

KW tissue typing; Bcr-abl; IFV; influenza.

KW

XX Mus sp.

OS

XX Key Location/Qualifiers

FT Misc-difference 12

FT /note= "biotinylated"

XX

FN WO9641188-A1.

XX

PD 19-DEC-1996.

XX

XX 07-JUN-1996; 96WO-US09680.

PF

XX 07-JUN-1995; 95US-0485610.

PR

XX (UNIW) UNIV WASHINGTON.

PA

XX Cheever MA, Chen W;

PI

XX WPI; 1997-108657/10.

DR

XX Identifying major histocompatibility complex class I binding mols. -

PT using peptide(s) having a core of 7-14 amino acids with extra amino

PT acids and a reporter gp. at the N- or C-terminus, useful for tissue

PT typing

XX

PS Example 3; Page 23; 41pp; English.

XX

CC AAW14122 is a biotinylated peptides derived from the OVA (ovalbumin)

CC

CC protein which can be obtained from either an LSTRA or EL-4 tumour

CC of Balb/c mice. The peptides bind to MHC class I molecules. This

CC is useful for tissue typing or for screening for molecules that

CC interact with MHC class I molecules. MHC class I molecules can be

CC identified using the peptides and also the peptides are useful in

CC vaccines against disease and infection e.g. caused by viruses,

CC bacteria or tumours.

XX

SQ Sequence 12 AA;

Query Match 100.0%; Score 38; DB 18; Length 12;

Best Local Similarity 100.0%; Pred. No. 0.18;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8

DB 1 SIINFEKL 8

RESULT 81

AAU09822

ID AAU09822 standard; peptide; 12 AA.

XX

AC AAU09822;

XX 14-FEB-2002 (first entry)

DT Modified ovalbumin-derived class I H-2Kb restricted peptide #2.

DE

XX Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;

KW immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;

KW popliteal lymph node; spleen; immune response; systemic response.

XX

OS Synthetic.

XX

XX WO200178767-A2.

PN

XX 25-OCT-2001.

PD

XX 17-APR-2001; 2001WO-EP04313.

PF

XX 14-APR-2000; 2000AT-0000657.

PR

XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.

PA

XX Mattner F, Zauner W, Schmidt W, Buschle M;

PI

XX WPI; 2002-025970/03.

DR

XX Pharmaceutical preparation for use as a potent vaccine for inducing an

PT improved immune response in a mammal, comprises a modified peptide -

PT

XX Example 1; Page 9; 18pp; English.

XX

CC The invention relates to a pharmaceutical preparation comprising a

CC modified peptide, which induces an improved immune response in a mammal

CC compared to the wild type peptide. The neutral peptide (SIINFEKL) (one

CC negatively charged (Glu), one positively charged (Lys) amino acid) was

CC rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu

CC Asp, respectively. Results showed that the addition of 4 negatively-

CC charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes

CC this peptide (in combination with poly-L-arginine) able to induce a high

CC amount of specific interferon (IFN)-gamma-producing T cells in the

CC (draining popliteal) lymph node (local response) and in the spleen

CC (systemic response). Thus, the addition of hydrophobic amino acids as

CC well as the addition of negatively charged amino acids transforms the

CC peptide SIINFEKL to a good inducer of specific T cells. The modified

CC peptides of the pharmaceutical composition induce a stronger immune

CC response in a mammal compared to wild type antigens. The present

CC sequence represents modified ovalbumin-derived class I H-2Kb restricted

CC peptide #2 as described in the method of the invention.

XX

SQ Sequence 12 AA;

! FINDPATTERNS on geneseq:* allowing 0 mismatches

! 1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {

AAB13781 ck: 1303 len: 25 ! Aab13781 Soluble peptide antigen pEA. 11/20

1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(E) (A) {5} (E) (A) {5} (E) (A) {5}

2: C EAAAAEAAAAEAAAAA EAAAA

(E) (A) {5} (E) (A) {5}

8: AAAAA EAAAAEAAAAEAAAAA

AAB13783 ck: 4553 len: 45 ! Aab13783 Soluble tandem pEA/ pK peptide con

1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(E) (A) {5} (E) (A) {5} (E) (A) {5}

2: C EAAAAEAAAAEAAAAA EAAAA

(E) (A) {5} (E) (A) {5}

8: AAAAA EAAAAEAAAAEAAAAA KKKK

ABP02760 ck: 5947 len: 86 ! Abp02760 Human ORFX protein sequence SEQ ID

1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(D) (A,G,C) {5} (E) (A,L,G,C,V) {5} (E) (A,G,C,V) {5}

13: GGRVR DACGCCVCGALEGAVCG LQEGP

Databases searched:

Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds: 5
Total length: 158,726,570
Total sequences: 1,107,863
CPU time: 09:16.70

```

!!SEQUENCE LIST 1.0
! FINDPATTERNS on geneseq:* allowing 0 mismatches
!      1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {
GENESEQP2000S:AAB13781 ck: 1303 len: 25 finds: 2 | Aab13781 Soluble peptide and
GENESEQP2000S:AAB13783 ck: 4553 len: 45 finds: 2 | Aab13783 Soluble tandem pEA/
GENESEQP2002S:ABP02760 ck: 5947 len: 86 finds: 1 | Abp02760 Human ORFX protein
\\End of list

```

```

Databases searched:
  Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003
Total finds: 5
Total length: 158,726,570
Total sequences: 1,107,863
CPU time: 11:32.04

```

```

!!SEQUENCE LIST 1.0
! FINDPATTERNS on geneseq.* allowing 0 mismatches
!      1 C(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (B,D) (A,L,I,F,G,C,M,V)
GENESQ2000S:AAB13781  ck: 1303  len: 25  finds: 1  ! Aab13781 Soluble peptide and
GENESQ2000S:AAB13783  ck: 4553  len: 45  finds: 1  ! Aab13783 Soluble tandem p2A/
\\End of list

Databases searched:
  Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds:      2
Total length:    158,726,570
Total sequences: 1,107,863
CPU time:        08:18.53

```


! FINDPATTERNS on pir:* allowing 0 mismatches

! 1 C(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V)

Databases searched:

NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

Total finds: 0

Total length: 96,168,682

Total sequences: 283,308

CPU time: 02:26.82

! FINDPATTERNS on swp:* allowing 0 mismatches

! 1 C(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V)

Databases searched:

SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

Total finds: 0
Total length: 305,079,309
Total sequences: 958,388
CPU time: 08:14.71


```
!!AA_SEQUENCE 1.0
ID AAB13781 standard; peptide; 25 AA.
XX AC AAB13781;
XX DT 10-NOV-2000 (first entry)
XX DE Soluble peptide antigen pEA.
XX KW pEA peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
XX KW major histocompatibility complex class I; MHC class I; antigen; tumour;
XX KW prostate; breast; multiple myeloma.
XX OS Unidentified.
XX PN WO200035949-A1.
XX PD 22-JUN-2000.
XX PF 14-DEC-1999; 99WO-US29724.
XX PR 14-DEC-1998; 98US-0112324.
XX PA (DEND-) DENDREON CORP.
XX PI Laus R, Hakim I, Vidovic D;
XX WPI; 2000-442365/38.
XX DR Antigen modified by the covalent addition of a peptide that
XX PT facilitates entry into antigen presenting cells, useful for producing
XX PT compositions for immunizing against tumors and pathogens -
XX PS Claim 2; Page 26; 34pp; English.
XX CC The present invention relates to compositions of modified soluble protein
XX CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
XX CC response i.e. a major histocompatibility complex (MHC) class I molecule
XX CC response. The protein antigen is modified by the covalent addition of a
XX CC peptide sequence which facilitate entry of the antigen into antigen
XX CC presenting cells (APCs). The present sequence is one such peptide
XX CC sequence which can be used to modify the soluble antigens. The present
XX CC sequence is tandem pEA/ pK peptide conjugate. The modified antigen
XX CC composition may be used for immunising against, or treating a tumour e.g.
XX CC prostate and breast carcinoma or multiple myeloma, or pathogen in
XX CC mammals.
XX SQ Sequence 25 AA;
AAB13781 Length: 25 January 30, 2004 11:00 Type: P Check: 1303 ..
!!AA_SEQUENCE 1.0
ID AAB13781 standard; peptide; 45 AA.
XX AC AAB13781;
XX DT 10-NOV-2000 (first entry)
XX DE Soluble tandem pEA/ pK peptide conjugate.
XX KW pK peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
XX KW major histocompatibility complex class I; MHC class I; antigen; tumour;
XX KW prostate; breast; multiple myeloma; pEA peptide.
XX OS Unidentified.
XX PN WO200035949-A1.
XX PD 22-JUN-2000.
XX PF 14-DEC-1999; 99WO-US29724.
XX PR 14-DEC-1998; 98US-0112324.
XX PA (DEND-) DENDREON CORP.
XX PI Laus R, Hakim I, Vidovic D;
XX WPI; 2000-442365/38.
XX DR Antigen modified by the covalent addition of a peptide that
XX PT facilitates entry into antigen presenting cells, useful for producing
XX PT compositions for immunizing against tumors and pathogens -
XX PS Claim 2; Page 26; 34pp; English.
XX CC The present invention relates to compositions of modified soluble protein
XX CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
XX CC response i.e. a major histocompatibility complex (MHC) class I molecule
XX CC response. The protein antigen is modified by the covalent addition of a
XX CC peptide sequence which facilitate entry of the antigen into antigen
XX CC presenting cells (APCs). The present sequence is one such peptide
XX CC sequence which can be used to modify the soluble antigens. The present
XX CC sequence is peptide pEA. The modified antigen composition may be used for
XX CC immunising against, or treating a tumour e.g. prostate and breast
XX CC carcinoma or multiple myeloma, or pathogen in mammals.
XX SQ Sequence 45 AA;
AAB13783 Length: 45 January 30, 2004 11:00 Type: P Check: 4553 ..
!!AA_SEQUENCE 1.0
ID AAB13783 standard; peptide; 86 AA.
XX AC AAB13783;
XX DT 25-JUN-2002 (first entry)
XX DE Human ORFX protein sequence SEQ ID NO:5502.
XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX KW hypertension; hypothyroidism; cholesterol ester storage disease;
XX KW immune deficiency; immune disorder; infectious disease;
XX KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX KW myasthenia gravis.
XX OS Homo sapiens.
XX PN WO200192523-A2.
XX PD 06-DEC-2001.
XX PF 29-MAY-2001; 2001WO-US10836.
XX PR 30-MAY-2000; 2000US-206132P.
XX PR 29-AUG-2000; 2000US-228716P.
XX PA (CURA-) CURAGEN CORP.
XX PI Shinkets RA, Leach MD;
XX WPI; 2002-106308/14.
XX N-PSDB; ABN18512.
XX DR Novel human polypeptides and polynucleotides useful for diagnosing,
XX KW preventing and treating cardiovascular disease, neurodegenerative,
XX KW hyperproliferative disorders and autoimmune disorders -
XX PS Disclosure; SEQ ID 5502; 1037pp; English.
```

XX The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIFO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 86 AA;

ABP02760 Length: 86 January 30, 2004 11:00 Type: P Check: 5947 ..

1 XPPTDCEGGR VRDAGCCCEV CGALEGAVCG LQEGPCGEGA ANAVSAPPFG

51 VPASATVRRR AQAGLCVCAS SEPVCGNDAK TYNLC

! FINDPATTERNS on swp:* allowing 0 mismatches

```

1 1 (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {
HRA1_MOUSE ck: 7206 len: 480 ! Q9r118 mus musculus (mouse). serine protease
(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(D) (A,G,C) {5} (E) (A,L,G,C,V) {5} (E) (A,G,C) {5}
60: GGRVR DACGCCVCGALEGACG LQEGP

Q9KHD9 ck: 1697 len: 262 ! Q9khd9 streptomyces griseus subsp. griseus
(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(E) (A,L,C) {5} (D) (A,L,G,V) {5} (E) (A,L,F,G) {5}
124: LGGGL ELALACDLVWAGEALFG LPELFG

Q8RUS6 ck: 3612 len: 107 ! Q8rus6 oryza sativa (japonica cultivar-group)
(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(E) (A,I,V) {5} (E) (G,C,V) {5} (D) (A,L,G,M,V) {5}
20: LETEV EAIAVREGGCGVDIAVVG RALGL

Q9QZK6 ck: 7594 len: 480 ! Q9qzk6 mus musculus (mouse). insulin-like growth factor
(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(D) (A,G,C) {5} (E) (A,L,G,C,V) {5} (E) (A,G,C) {5}
60: GGRVR DACGCCVCGALEGACG LQEGP

Q91WS3 ck: 7882 len: 480 ! Q91ws3 mus musculus (mouse). protease, serine
(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(D) (A,G,C) {5} (E) (A,L,G,C,V) {5} (E) (A,G,C) {5}
60: GGRVR DACGCCVCGALEGACG LQEGP

Q9QZK5 ck: 8689 len: 480 ! Q9qzk5 rattus norvegicus (rat). insulin-like growth factor
(E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,F,G,C,M,V) {5,5} (E,D) (A,L,I,
(D) (A,G,C) {5} (E) (A,L,G,C,V) {5} (E) (A,G,C,V) {5}
60: GGRVR DACGCCVCGALEGACG LQEGP

```

Databases searched:

SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
 SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

Total finds: 6

Total length: 305,079,309

Total sequences: 958,388

CPU time: 14:58.50

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 07:06:27 ; Search time 70 Seconds
(without alignments)
29.492 Million cell updates/sec

Title: SEQ10

Perfect score: 38

Sequence: 1 siinfekl 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPREMBL23.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	86.8	168	10 Q98RQ3	Q98rq3 guillardia
2	33	86.8	315	10 O65158	O65158 phaseolus v
3	33	86.8	366	2 Q8VQ88	Q8vq88 uncultured
4	32	84.2	260	16 Q9AA01	Q9aa01 caulobacter
5	32	84.2	299	16 Q92XR1	Q92xr1 rhizobium m
6	32	84.2	347	8 Q98RV6	Q98rv6 guillardia
7	32	84.2	443	17 Q8TRP5	Q8trp5 methanosarc
8	32	84.2	444	4 Q8IXU5	Q8ixu5 homo sapien
9	32	84.2	580	4 Q92565	Q92565 homo sapien
10	32	84.2	612	11 Q8BJJ9	Q8bjj9 mus musculus
11	32	84.2	814	11 Q8CORS	Q8cor5 mus musculus
12	32	84.2	814	11 Q8CQ99	Q8cq99 mus musculus
13	32	84.2	965	2 Q9S4D1	Q9s4d1 staphylococ
14	32	84.2	965	2 Q8VVR0	Q8vvr0 staphylococ
15	32	84.2	1311	5 Q95NL8	Q95nl8 caenorhabdi
16	32	84.2	1347	5 Q95WR8	Q95wr8 caenorhabdi

17	32	84.2	1470	5 Q21218	Q21218 caenorhabdi
18	32	84.2	1573	5 Q9VMF3	Q9vmf3 drosophila
19	32	84.2	1573	5 Q95V18	Q95v18 drosophila
20	31	81.6	97	12 Q8BDR4	Q8bdr4 reindeer pa
21	31	81.6	103	5 Q81B53	Q81b53 plasmodium
22	31	81.6	211	5 P91370	P91370 caenorhabdi
23	31	81.6	287	11 Q8R1R1	Q8r1r1 mus musculus
24	31	81.6	307	17 O59582	O59582 pyrococcus
25	31	81.6	309	16 Q98QRS	Q98qr5 mycoplasma
26	31	81.6	364	10 Q98KB4	Q98kb4 arabidopsis
27	31	81.6	461	11 Q8BZK9	Q8bzkr9 mus musculus
28	31	81.6	680	12 Q9YTO6	Q9ytc6 ateline her
29	31	81.6	834	11 Q8R3E5	Q8r3e5 mus musculus
30	31	81.6	876	11 Q8VCC8	Q8vcc8 mus musculus
31	31	81.6	881	4 Q8WVN0	Q8wvn0 homo sapien
32	31	81.6	881	4 Q95634	Q95634 homo sapien
33	31	81.6	881	4 Q95398	Q95398 homo sapien
34	31	81.6	884	11 Q921C8	Q921c8 rattus norv
35	31	81.6	1089	5 Q81605	Q81605 plasmodium
36	31	81.6	1113	4 Q8TEA3	Q8tea3 homo sapien
37	31	81.6	1138	11 Q8CHG7	Q8chg7 mus musculus
38	31	81.6	1204	4 Q9UHV4	Q9uhv4 homo sapien
39	31	81.6	1391	4 Q8TEU6	Q8teu6 homo sapien
40	31	81.6	1499	4 Q9Y4G8	Q9y4g8 homo sapien
41	31	81.6	1509	4 Q96FC1	Q96pci homo sapien
42	31	81.6	1601	4 Q8TEU7	Q8teu7 homo sapien
43	31	81.6	1601	4 Q8NI21	Q8ni21 homo sapien
44	31	81.6	1836	10 Q9LXR4	Q9lrx4 arabidopsis
45	31	81.6	1909	10 Q9LXR3	Q9lrx3 arabidopsis

ALIGNMENTS

RESULT 1

Q98RQ3 PRELIMINARY; PRT; 168 AA.
AC Q98RQ3
DT 01-OCT-2001 (TREMELrel. 18, Created)
DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMELrel. 18, Last annotation update)
DE Hypothetical 20.2 kDa protein orf168 from chromosome 1.
GN ORF168.
OS Guillardia theta (Cryptomonas phi).
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2122349; PubMed=11323671;
RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.G.;
RT "The highly reduced genome of an enslaved algal nucleus."
RL Nature 410.1091-1096(2001).
DR EMBL, AF165818; AAK39893.1; --
KW Hypothetical protein.
SQ SEQUENCE 168 AA; 20185 MW; 2874CBD53028A3DD CRC64;

Query Match 86.8%; Score 33; DB 10; Length 168;
Best Local Similarity 75.0%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEL 8
DB 108 NIINFEL 115

RESULT 2

O65158 PRELIMINARY; PRT; 315 AA.
ID O65158
AC O65158;
DT 01-AUG-1998 (TREMELrel. 07, Created)
DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)

```

DE 1-aminocyclopropane-1-carboxylic acid oxidase.
GN ACO1.
OS Phaseolus vulgaris (kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Taylor; TISSUE=Leaf;
RX MEDLINE=21173698; PubMed=11259647;
RA Pidgeon C.M., Reid D.M., Facchini P.J.;
RT "Light induced changes in ethylene production in Phaseolus vulgaris
cv. Taylor.";
RL Plant Physiol. 114:165-165 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Taylor; TISSUE=Leaf;
RX Pidgeon C.M., Facchini P.J., Reid D.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF053354; AAC12934.1; -.
DR InterPro; IPR005123; 2OG-FerII_Oxy.
DR Pfam; PF03171; 2OG-FerII_Oxy; 1.
SQ SEQUENCE 315 AA; 36141 MW; 6898811DF98E9B81 CRC64;

Query Match      86.8%; Score 33; DB 10; Length 315;
Best Local Similarity 85.7%; Pred. No. 72;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IINFEKL 8
Db :|||||
6 VINFEKL 12

RESULT 3
Q8VQ88 PRELIMINARY; PRT; 366 AA.
AC Q8VQ88;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Formyltetrahydrofolate synthetase (Fragment).
DE uncultured environmental landfill bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=184235;
RN [1]
RP SEQUENCE FROM N.A.
RA Uz I., Lindner A.S., Rasche M.E., Townsend T.G., Ogram A.V.;
RT "Microbial Characterization of Landfill Cover Soils of Different
Ages.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF459462; AAL67505.1; -.
DR InterPro; IPR000559; Fmtethyd_synth.
DR Pfam; PF01268; FTHFS; 1.
FT NON_TER 1
FT NON_TER 366
SQ SEQUENCE 366 AA; 39816 MW; EB4685D501E97026 CRC64;

Query Match      86.8%; Score 33; DB 2; Length 366;
Best Local Similarity 85.7%; Pred. No. 83;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IINFEKL 8
Db :|||||
346 IINFEKL 352

RESULT 4
Q9AA01 PRELIMINARY; PRT; 260 AA.
AC Q9AA01;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

```

```

DE Hypothetical protein CC0810.
GN CC0810.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Ely B.,
RA Fockea I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Shetty K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).
DR EMBL; AB005757; AAK22795.1; -.
DR TIGR; CC0810; -.
DR InterPro; IPR005269; Cons_hypoth730.
DR Pfam; PF03641; Lysine decarbox; 1.
DR TIGRFAMs; TIGR00730; TIGR00730; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 260 AA; 28789 MW; E4CF85223B12EF1B CRC64;

Query Match      84.2%; Score 32; DB 16; Length 260;
Best Local Similarity 87.5%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
Db :|||||
215 SIINFEAL 222

RESULT 5
Q92KR1 PRELIMINARY; PRT; 299 AA.
AC Q92KR1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein R00434.
GN R00434 OR SMC01731.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
DR EMBL; AL591783; CAC41871.1; -.
DR InterPro; IPR005269; Cons_hypoth730.
DR Pfam; PF03641; Lysine decarbox; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 299 AA; 33440 MW; 9BD61375BE91F195 CRC64;

Query Match      84.2%; Score 32; DB 16; Length 299;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
Db :|||||

```

```

Db      246 SIINFEAL 253

RESULT 6
Q98RV6 PRELIMINARY; PRT; 347 AA.
ID Q98RV6
AC Q98RV6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Cyclin B.
GN CYCB.
OS Guillardia theta (Cryptomonas phi).
OG Nucleomorph.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21223349; PubMed=11323671;
RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.G.;
RT "The highly reduced genome of an enslaved algal nucleus.";
RL Nature 410:1091-1096 (2001).
CC -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY.
DR EMBL; AF165818; AAK39844.1; --
DR InterPro; IPR006670; Cyclin.
DR InterPro; IPR006671; Cyclin_N.
DR Pfam; PF00134; cyclin; 1.
DR SMART; SMO0385; CYCLIN; 2.
DR PROSITE; PS00292; CYCLINS; 1.
DR Cell cycle; Cell division; Cyclin.
KW Cell cycle; Cell division; Cyclin.
SQ SEQUENCE 347 AA; 41616 MW; GC579909C403A04D CRC64;

Query Match 84.2%; Score 32; DB 8; Length 347;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
DB 86 NLNFEKL 93

RESULT 7
Q8TRP5 PRELIMINARY; PRT; 443 AA.
ID Q8TRP5
AC Q8TRP5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Iron ABC transporter, solute-binding protein.
GN MA1130.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CZA / ATCC 35395 / DSM 2834;
RA MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atncor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.B., Graham D.A., Guss A.M.,
RA Hederich R., Ingram-Smith C., Kuettnar H.C., Krazycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam I.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542 (2002).

DR EMBL; AE010779; AAM04551.1; --
DR InterPro; IPR002491; Peripla_BP.
DR Pfam; PF01497; Peripla_BP_2; 1.
KW Complete proteome.
SQ SEQUENCE 443 AA; 49537 MW; 0D3F2198B20E88C9 CRC64;

Query Match 84.2%; Score 32; DB 17; Length 443;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
DB 166 TIINFEKL 173

RESULT 8
Q8IXU5 PRELIMINARY; PRT; 444 AA.
ID Q8IXU5
AC Q8IXU5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to guanine nucleotide exchange factor for Rap1,
DE M-Ras-regulated GEF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX Strausberg R.;
RC TISSUE=Brain;
SQ SEQUENCE 444 AA; 52051 MW; ECBEDF820DDFF13B CRC64;
DR EMBL; BC039203; AAH39203.1; --
DR Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
SQ SEQUENCE 444 AA; 52051 MW; ECBEDF820DDFF13B CRC64;

Query Match 84.2%; Score 32; DB 4; Length 444;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
DB 381 NLNFEKL 388

RESULT 9
Q92565 PRELIMINARY; PRT; 580 AA.
ID Q92565
AC Q92565;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein KIAA0277.
GN KIAA0277.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RA MEDLINE=97151544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329 (1996).
DR EMBL; D87467; BAAL3406.1; --
DR InterPro; IPR00651; RasGEFN.
DR IPfam; IPR001895; RasGEFN.
DR Pfam; PF00617; RasGEF; 1.
DR Pfam; PF00618; RasGEFN; 1.

```

```
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 1.
KW Hypothetical protein.
SQ SEQUENCE 580 AA; 67733 MW; 732FB/AA11DFDA1C CRC64;

Query Match      84.2%; Score 32; DB 4; Length 580;
Best Local Similarity 62.5%; Pred. No. 2.1e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
   :::|||||
Db 517 NLVNFEL 524

RESULT 10
Q8BJJ9 PRELIMINARY; PRT; 612 AA.
AC Q8BJJ9;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical guanine-nucleotide dissociation stimulators CDC25
DE family/guanine nucleotide exchange factor for Ras-like GTPases;
DE N-terminal motif containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK083591; BAC38963.1; -.
KW Hypothetical protein.
SQ SEQUENCE 612 AA; 70913 MW; DBD6552DDFD392B6 CRC64;

Query Match      84.2%; Score 32; DB 11; Length 612;
Best Local Similarity 62.5%; Pred. No. 2.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
   :::|||||
Db 548 NLVNFEL 555

RESULT 11
Q8C0R5 PRELIMINARY; PRT; 814 AA.
AC Q8C0R5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical guanine-nucleotide dissociation stimulators CDC25
DE family/guanine nucleotide exchange factor for Ras-like GTPases;
DE N-terminal motif containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).

DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 1.
KW Hypothetical protein.
SQ SEQUENCE 580 AA; 67733 MW; 732FB/AA11DFDA1C CRC64;

Query Match      84.2%; Score 32; DB 11; Length 814;
Best Local Similarity 62.5%; Pred. No. 2.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
   :::|||||
Db 750 NLVNFEL 757

RESULT 12
Q8C0Q9 PRELIMINARY; PRT; 814 AA.
AC Q8C0Q9;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical guanine-nucleotide dissociation stimulators CDC25
DE family/guanine nucleotide exchange factor for Ras-like GTPases;
DE N-terminal motif containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK030016; BAC26736.1; -.
KW Hypothetical protein.
SQ SEQUENCE 814 AA; 93754 MW; AA04EEC74CE13735 CRC64;

Query Match      84.2%; Score 32; DB 11; Length 814;
Best Local Similarity 62.5%; Pred. No. 2.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEKL 8
   :::|||||
Db 750 NLVNFEL 757

RESULT 13
Q9S4D1 PRELIMINARY; PRT; 965 AA.
AC Q9S4D1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Lantibiotic modifying enzyme.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C55;
RX MEDLINE=99346225; PubMed=10417203;
RA Navaratna M.A., Sahi H.G., Tagg J.R.;
RT "Identification of genes encoding two-component lantibiotic production
RT in Staphylococcus aureus C55 and other phage group II S. aureus
RT strains and demonstration of an association with the exfoliative toxin
RT B gene.";
RL Infect. Immun. 67:4268-4271 (1999).
DR EMBL; AF147744; AAD47013.1; -.
SQ SEQUENCE 965 AA; 111570 MW; BC1E4DABDAC4F346 CRC64;
```

Query Match 84.2%; Score 32; DB 2; Length 965;
Best Local Similarity 71.4%; Pred. No. 3.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEX 7
Db 202 SVNFEK 208
[1]:::|||||

RESULT 14
Q8VVR0 PRELIMINARY; PRT; 965 AA.
AC Q8VVR0;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE ORF42.
OS Staphylococcus aureus.
OG Plasmid ETB plasmid.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TV4;
RX MEDLINE=21562640; PubMed=11705958;
RA Yamauchi T., Hayashi T., Takami H., Ohnisi M., Murata T.,
RA Nakayama K., Asakawa K., Ohara M., Komatsuzawa H., Sugai M.,
RT "Complete Nucleotide Sequence of a Staphylococcus aureus Exfoliative
RT Toxin B Plasmid and Identification of a Novel ADP-Ribosyltransferase,
RT EDIN-C";
RL Infect. Immun. 69:7760-7771(2001).
DR ENBL; AP003088; BAB78440.1; -.
KW Plasmid.
SQ SEQUENCE 965 AA; 111561 MW; C08E9FD8DAC5F82B CRC64;

Query Match 84.2%; Score 32; DB 2; Length 965;
Best Local Similarity 71.4%; Pred. No. 3.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIINFEX 7
Db 202 SVNFEK 208
[1]:::|||||

RESULT 15
Q95NL8 PRELIMINARY; PRT; 1311 AA.
AC Q95NL8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE T14G10.2b protein (PXF isoform B).
GN T14G10.2 OR T14G10.2B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wild A.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Verheijen M., van Berkel W., Jansen G., de Rooij J., Plasterk R.H.,
RA Bos J.L., Zwartkruis F.J.T.;
RT "Characterization of pxf, the C. elegans homolog of human PDZ-GEFs.";

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Wild A.;
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR ENBL; Z68880; CAC42342.1; -.
DR ENBL; Z69664; CAC42342.1; JOINED.
DR ENBL; AF308448; AAL09434.1; -.
DR ENBL; Z68880; CAC42313.1; -.
DR WormPep; T14G10.2b; CE28081.
DR InterPro; IPR000595; CNMP_binding.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR000651; RasGEFN.
DR InterPro; IPR001895; RasGRF_CDC25.
DR InterPro; IPR000159; RA domain.
DR Pfam; PF00027; CNMP_binding; 2.
DR Pfam; PF00595; PDZ_1.
DR Pfam; PF00788; RA; 1.
DR Pfam; PF00617; RasGEF; 1.
DR SMART; SM00100; CNMP; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00314; RA; 1.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 1.
DR PROSITE; P550042; CNMP_BINDING_3; 1.
DR PROSITE; P550106; PDZ; 1.
SQ SEQUENCE 1311 AA; 147003 MW; A4870F07DC201F97 CRC64;

Query Match 84.2%; Score 32; DB 5; Length 1311;
Best Local Similarity 85.7%; Pred. No. 4.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IINFEXL 8
Db 1068 IINFEXL 1074
[1]:::|||||

Search completed: January 30, 2004, 07:09:11
Job time : 71 secs


```

RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RN Nature 420:563-573(2002).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[4]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[5]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
Kanno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[6]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076661;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
Kanno H., Akiyama J., Nishii K., Kitzunai T., Tashiro H., Itoh M.,
Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
Yanamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
DR EMBL; AK015384; BAB29822.2; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 169 AA; 19305 MW; 91B9959380A694CC CRC64;

Q9D5G1 Length: 169 January 30, 2004 08:18 Type: P Check: 9388

1 RCTGQAGPOL RALAGPGWPR LAPALLSGGR ARNIAGLPAA KHAPESGAS

51 ARRLPAAPH RGGQGDAA SRELASTHG RNLPHCLPLP PLAQTSSLPV

101 WYRHLKPKIS SVQLGRRRR RRRRRRRR RRRRRRRK KEEEEEVVA

151 FLDSLIEPLR TSNLHRHE

!!AA SEQUENCE 1.0
ID Q35807 PRELIMINARY; PRT; 129 AA.
AC Q35807;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE MICROVASCULAR endothelial differentiation protein 2.
GN MDG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Epithelium;
RX MEDLINE=98172708; PubMed=9511718;
RA Proels F., Loser B., Marx M.;
RT "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA

```

```

RT species, during in vitro angiogenesis.";
RL Exp. Cell Res. 239:1-10(1998).
DR EMBL; Y08769; CAA70022.1; -.
DR InterPro; IPR000719; Prot kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot kinase; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 129 AA; 15080 MW; 38102272BBE2EDB4 CRC64;

O35807 Length: 129 January 30, 2004 08:18 Type: P Check: 7510

1 MKKPHIVEL LETYSSDGL YVFEFMDGA DLCFEIVRA DAGFVYSEAV
51 ASHYMRQILE ALRYCHDNNI IHRDVKPHCV LIASKKKKK KKKKKKKKK
101 KKKKKIKWEG RDAFWAIPV KSSGGVLIQ

!!AA SEQUENCE 1.0
ID Q8BXG9 PRELIMINARY; PRT; 115 AA.
AC Q8BXG9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical arginine-rich region containing protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK047167; BAC32979.1; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 115 AA; 13938 MW; 6C0F7EED8669CF65 CRC64;

Q8BXG9 Length: 115 January 30, 2004 08:18 Type: P Check: 5434

1 GRGRRISEFE VSQYTEKPC LEKPKKKKK I11111IIR RRRRRRRR

51 RRRRRRKK ETGSHFVALA SLELHLPPEC WNLRYRLPHP TQIGFNAIKI

101 SIKOTSLSLL AFCVK

!!AA SEQUENCE 1.0
ID Q8BHV2 PRELIMINARY; PRT; 154 AA.
AC Q8BHV2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Weakly similar to hypothetical 10.3 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK082253; BAC38447.1; -.
KW Hypothetical protein.

```

!!AA_SEQUENCE 1.0
P1:T49173 - hypothetical protein T20N10.250 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
C:Accession: T49173
R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.;
Rudd, S.; Lemcke, K.; Mayer, K.P.X.; Quetier, F.; Salanoubat, M.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25017
A:Status: preliminary
A:Accession: T49173
A:Molecule type: DNA
A:Residues: 1-517 <DAN>
A:Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250
A:Experimental source: cultivar Columbia; BAC clone T20N10
C:Genetics:
A:Gene: ATSP:T20N10.250
A:Map position: 3
A:Introns: 312/3; 359/3; 444/3
C:Superfamily: Arabidopsis thaliana hypothetical protein P17J16.30

T49173 Length: 517 January 30, 2004 07:58 Type: P Check: 4143 ..
1 MDLFSPLNE LLYHLSPLS TKEAALTSVL SKRWNLPAF VPYLEFDDSV
51 FLHPERKRE KEGILQSPMD FVDRVLDLHG DSLIKTFSLK CKTGVDSDHV
101 DRWICNVLAR GVSOLDLFD FRLYSLPHE VGVSRVLVL RVGSESDLYW
151 WQKFLCLPML KTLVLDSCWL CIGQFOILL ACPALEELDM TNTRWKSINV
201 TVSSSILKEL TIDHGGCSV VNLKLSFDA PSLVTFYYCD SLAEDYQOVN
251 LKNLYEAQIN LLLTQAQIEQ VRALNNEMLV ADDVFPFGLN AWKLITGLRN
301 VQQLYLSPTD LEVLSRCEG MPVFNLLKVL SIWSDMNRGM QAMPVLLRNC
351 PHLEWLIIEG LLHYATDKCG DVDCISRDY KOHSITSCPV KKLQIYFRRG
401 TIRELEMIXH FLKIFPCLKE MDIYAHENSH TLFKDPTIFE RVGKKKKKKK
451 KKKKKKKKK KKKKKKIRLN FKPVNTEQF LKRLADKLCF IPOCLEFLDV
501 DSSLGELALL AMDSRPS

!!AA_SEQUENCE 1.0
P1:T46395 - hypothetical protein DKFPz434I1120.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46395
R:Ottewälder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23031
A:Accession: T46395
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-380 <AAA>
A:Cross-references: EMBL:AL137556
A:Experimental source: adult testis; clone DKFPz434I1120
C:Genetics:
A:Note: DKFPz434I1120.1

T46395 Length: 380 January 30, 2004 07:58 Type: P Check: 7330 ..
1 MGSTDSKLNK RKAVIQLTTK TQPVEATDDA FWDQFWADTA TSVDQVFPALV
51 PAAEIRAVRE ESPNLATLC YKAVEKLVOG AESGCHSEKE KQIVLNC SRL
101 LTRVLPIYFE DPDWRGFWMS TVPAGRGGG EEDDEHARPL AESLLLLAIAD
151 LLFCPDFTVQ SHRRSTVDSA EDVHSLDSC EYIWEAGVGFA HSPQPNVIHD
201 MNRMLLKLK LTCSEAMYL PPAESGSTN PWVQFCSTE NRHALPLFTS

251 LLNTVCAYDP VGYGIPYNHL LFSDYREPLV EAAQVLIVT LDHDSASSAS
301 PTVDTTGT AMDDADPPGP ENLFVNYLSR IHRBEDFOFI LKGIARLLSN
351 LLLQKKKKKK KKKKKKKKK KKKKKKKKK

!!AA_SEQUENCE 1.0
P1:I52523 - nucleoporin p62 homolog - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 05-Nov-1999
C:Accession: I52523
R:Wang, Z.Q.; Akmal, K.M.; Kim, K.H.
Bio. Reprod. 51, 1022-1030, 1994
A:Title: An unusual nucleoporin-related messenger ribonucleic acid is present
in the germ cells of rat testis.
A:Reference number: I52523; MUID:95151924; PMID:7849178
A:Accession: I52523
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-215 <RES>
A:Cross-references: GB:S75997; NID:G913245; PIDN:AB33384.1; PID:G913246
A:Experimental source: testis

I52523 Length: 215 January 30, 2004 07:58 Type: P Check: 8048 ..
1 SSGRATSSCD EDCSSSLPF SLSGPVKQDC EFLEKKKKKK KKKKKKKKKK
51 KKKKKKTGDN AKSVSRQYSL KVTLEHEAE QAKVELDFIL SQKKELDLL
101 SPLEESVKEQ SGTIYLOHAD EREKTYKLA ENIDAQLKRM AQDLKDIIIEH
151 LNMAGGPADT SDPLOQICKI LNAHMDSLQW VDQSSALLQR RVBEASRVCE
201 SRRKEQERSL RIAFD

!!AA_SEQUENCE 1.0
P1:S58321 - probable membrane protein YOR309c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O6105
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 19-Apr-2002
C:Accession: S58321; S67215; S71989
R:Pearson, B.M.; Hernando, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.
submitted to the EMBL Data Library, August 1995
A:Reference number: S58318
A:Accession: S58321
A:Molecule type: DNA
A:Residues: 1-126 <PEA>
A:Cross-references: EMBL:X90565; NID:G940836; PID:G940840
R:Pearson, B.M.; Hernando, Y.; Kalogeropoulos, A.; Schweizer, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67213
A:Accession: S67215
A:Molecule type: DNA
A:Residues: 1-126 <PEW>
A:Cross-references: EMBL:Z75217; NID:G1420680; PID:e252431; PID:G1420681;
MIPS:YOR309C
A:Experimental source: strain S288C
R:Pearson, B.M.; Hernando, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.;
Schweizer, M.
Yeast 12, 1021-1031, 1996
A:Title: Sequencing of a 35.71 kb DNA segment on the right arm of yeast
chromosome XV reveals regions of similarity to chromosomes I and XIII.
A:Reference number: S71986; MUID:97051589; PMID:8896266
A:Accession: S71989
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-126 <PEF>
A:Cross-references: EMBL:X90565; NID:G940836; PIDN:CAA62164.1; PID:G940840
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August
1995
C:Genetics:
A:Cross-references: SGD:S0005836

```
!!SEQUENCE LIST 1.0
! FINDPATTERNS on pir:* allowing 0 mismatches
1      1 (R,K){20,20}      January 30, 2004 07:51 ..
PIR2:T49173      ck: 4143 len: 517 finds: 4 | hypothetical protein T20N10.25
PIR2:T46395      ck: 7330 len: 380 finds: 7 | hypothetical protein DKF2p434t
PIR2:I52523      ck: 8048 len: 215 finds: 3 | nucleoporin p62 homolog - rat
PIR2:S58321      ck: 1384 len: 126 finds: 3 | probable membrane protein YOR3
```

\\End of list

Databases searched:
NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

Total finds: 17
Total length: 96,168,682
Total sequences: 283,308
CPU time: 03:53.65

! FINDPATTERNS on pir:* allowing 0 mismatches

! 1 (R,K){20,20}

January 30, 2004 07:05 ..

T49173 ck: 4143 len: 517 ! hypothetical protein T20N10.250 - Arabidops

(R,K){20,20}
(K){20}

444: FERVG KKKKKKKKKKKKKKKKKKKKKKK KKKIR

(K){20}
445: ERVCK KKKKKKKKKKKKKKKKKKKKK KIRL

(K){20}
446: RVGKK KKKKKKKKKKKKKKKKKKKKK KIRLN

(K){20}
447: VGKKK KKKKKKKKKKKKKKKKKKKKK IRLNF

T46395 ck: 7330 len: 380 ! hypothetical protein DKFZp434I1120.1 - huma

(R,K){20,20}
(K){20}

355: NLLLQ KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
356: LLLQK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
357: LLQKK KKKKKKKKKKKKKKKKKKKKK KKKK

(K){20}
358: LQKKK KKKKKKKKKKKKKKKKKKKKK KKK

(K){20}
359: QKKKK KKKKKKKKKKKKKKKKKKKKK KK

(K){20}
360: KKKKK KKKKKKKKKKKKKKKKKKKKK K

(K){20}
361: KKKKK KKKKKKKKKKKKKKKKKKKKK

I52523 ck: 8048 len: 215 ! nucleoporin p62 homolog - rat (fragment)

(R,K){20,20}
(K){20}

35: CEPLE KKKKKKKKKKKKKKKKKKKKK KKTGD

(K){20}
36: EFLEK KKKKKKKKKKKKKKKKKKKKK KTGDN

(K){20}
37: FLEKK KKKKKKKKKKKKKKKKKKKKK TGDNA

S58321 ck: 1384 len: 126 ! probable membrane protein YOR309c - yeast

(R,K){20,20}
(R,K){20}

53: RKERT RKKKKKKKKKKKKKKKKKKKK KRSPR

(R,K){20}
54: KRTR RKKKKKKKKKKKKKKKKKKKK RSPRK

(R,K){20}
55: RRTTR RKKKKKKKKKKKKKKKKKKKK SPRKR

Databases searched:

NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

Total finds: 17
Total length: 96,168,682
Total sequences: 283,308
CPU time: 03:26.44

! FINDPATTERNS on swp:* allowing 0 mismatches

! 1 C(R,K){20,20} January 30, 2004 06:58 ..

Databases searched:
SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

Total finds: 0
Total length: 305,079,309
Total sequences: 958,388
CPU time: 06:50.54

! FINDPATTERNS on pir:* allowing 0 mismatches

! 1 C(R,K){20,20} January 30, 2004 06:58 ..

Databases searched:

NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

Total finds: 0

Total length: 96,168,682

Total sequences: 283,308

CPU time: 01:56.42

CC disease). The polynucleotide sequences of the invention may also be
CC used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding
CC proteins.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 58 AA;

AAU18238 Length: 58 January 30, 2004 07:48 Type: P Check: 5509 ..

1 TYLECEHNSL VNSKCLTVL SRCISVCLNK FYFVCKKKKK KKKKKKKKK

51 KKKKKKKK

!!AA SEQUENCE 1.0
ID _AAO03766 standard; Protein; 81 AA.

XX AAO03766;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 17658.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-514838/56.

XX DR N-PSDB; AA183697.

XX PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -

XX PS Claim 20; SEQ ID NO 17658; 1399pp + Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 81 AA;

AAO03766 Length: 81 .January 30, 2004 07:48 Type: P Check: 8808 ..

1 GLNQTLRKI LAYSSITHG XIITAVLPYNP NITILNLTII IILTTACKK

51 KKKKKKKKKK KKKKKKKKKK KKKKKGGGG A

!!AA SEQUENCE 1.0

ID _AAO11210 standard; Protein; 70 AA.

XX AAO11210;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 25102.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-514838/56.

XX DR N-PSDB; AA191141.

XX PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -

XX PS Claim 20; SEQ ID NO 25102; 1399pp + Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 70 AA;

AAO11210 Length: 70 January 30, 2004 07:48 Type: P Check: 863 ..

1 YYIHRITVFM CMNXLKDNV DKXTIDLCLC KKKKKKKKK KKKKKKKKK

51 KKKKKKKKKK KKKKPOGGGA

PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225216.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246533.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-02559678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465557/50.

N-ESDB; AAS29114.

Nucleic acid molecules encoding human secreted chromosomal binding proteins, used in preventing treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -

Claim 11; SEQ ID No 223; 561pp; English.

The present invention relates to the isolation of novel DNA-binding proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding for these proteins. DNA-binding proteins such as histones, chromo (chromatin organisation modifier) domain proteins, and Y-box binding proteins may contribute to diseases resulting from aberrant DNA organisation and/or gene transcription. The sequences of the invention are useful in screening assays to identify antagonists and/or agonists that may enhance or block activities mediated by DNA-binding proteins. Blockers of DNA-binding proteins may be useful in treating disorders such as malignant diseases (e.g. cancer), autoimmune disorders (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's


```

!!AA SEQUENCE 1.0
ID AAW45801 standard; peptide; 39 AA.
XX
AC AAW45801;
XX
DT 25-JUN-1998 (first entry)
XX
DE One chain of a bombesin dimer.
XX
KW Alpha-melanocyte stimulating hormone; alpha-MSH; receptor agonist;
KW alpha-MSH-ANT; bombesin; dimer; bivalent agonist; disulphide bond;
KW G-protein coupled receptor.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Cross-links 1
FT /note= "This residue is disulphide bonded to the
FT corresponding Cys residue of an identical chain
FT to form a dimer"
FT Modified-site 28
FT /note= "Epsilon-aminohexanoic acid"
FT
XX WO9803632-A1.
XX
XX 29-JAN-1998.
XX
XX 23-JUL-1997; 97WO-US12911.
XX
XX 24-JUL-1996; 96US-0686934.
XX
XX (UYVA ) UNIV YALE.
XX
XX Carrithers MD, Lerner MR;
XX
XX WPI; 1998-120757/11.
XX
XX Bivalent agonist of G-protein coupled receptors containing two
XX ligand domains - bonded to molecular backbone, for treatment of
XX hypotension, promotion of skin tanning etc., also for delivering
XX drugs and gene therapy vectors to selected cells
XX
XX Claim 41; Page 48; 71pp; English.
XX
XX This sequence represents one of two identical chains disulphide bonded
XX to form a bombesin dimer. The invention relates to bivalent agonists,
XX with affinity for at least 1 G-protein coupled receptor (GPCR). The
XX bivalent agonists comprise: (a) two ligand domains (LD), individually
XX agonists or antagonists for GPCR, spaced 40-250 Angstrom apart, and (b)
XX a molecular backbone (MB) covalently bound to LD. The bivalent agonists
XX are useful in human or veterinary medicine as carriers for drugs or gene
XX therapy vectors, allowing these to be endocytosed by GPCR-expressing
XX cells. They can also be used e.g. to treat hypertension (angiotensin-
XX based LD); to increase levels of luteinising hormone (LH), using LD
XX derived from LH-releasing hormone, or to promote skin tanning (LD based
XX on alpha -melanocyte-stimulating hormone, MSH). The bivalent agonists
XX are administered orally, by injection or topically. Typical doses for
XX skin tanning are 1-4000 (especially 30-100) mu mole/kg systemically or
XX the bivalent agonists are administered topically in a composition
XX containing 0.001-10 (especially 1) mM. Where both LD are agonists, the
XX bivalent agonist has a synergistically higher activity than two
XX individual agonist ligands, and where at least 1 is an antagonist the
XX effect is stimulatory. The bivalent agonists are active at lower
XX concentrations than known agonists so should avoid toxicity problems.
XX
XX Sequence 39 AA;
XX
AAW45801 Length: 39 January 30, 2004 07:48 Type: P Check: 9500
XX
!!AA SEQUENCE 1.0
ID AAB13780 standard; peptide; 21 AA.
XX
1 CKKKKKKKKK KKKKKKKKK KXGGGQORLG NQWAVGHLM
XX

```

```

XX
AC AAB13780;
XX
DT 10-NOV-2000 (first entry)
XX
DE Soluble peptide antigen pk.
XX
KW pk peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW major histocompatibility complex class I; MHC class I; antigen; tumour;
KW prostate; breast; multiple myeloma.
XX
OS Unidentified.
XX
XX WO200035949-A1.
XX
XX 22-JUN-2000.
XX
XX 14-DEC-1999; 99WO-US29724.
XX
XX 14-DEC-1998; 98US-0112324.
XX
XX (DEND-) DENDREON CORP.
XX
XX Laus R, Hakim I, Vidovic D;
XX
XX WPI; 2000-442365/38.
XX
XX Antigens modified by the covalent addition of a peptide that
XX facilitates entry into antigen presenting cells, useful for producing
XX compositions for immunizing against tumors and pathogens -
XX
XX Claim 2; Page 26; 34pp; English.
XX
XX The present invention relates to compositions of modified soluble protein
XX antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
XX response i.e. a major histocompatibility complex (MHC) class I molecule
XX response. The protein antigen is modified by the covalent addition of a
XX peptide sequence which facilitate entry of the antigen into antigen
XX presenting cells (APCs). The present sequence is one such peptide
XX sequence which can be used to modify the soluble antigens. The present
XX sequence is peptide pk. The modified antigen composition may be used for
XX immunising against, or treating a tumour e.g. prostate and breast
XX carcinoma or multiple myeloma, or pathogen in mammals.
XX
XX Sequence 21 AA;
XX
AAB13780 Length: 21 January 30, 2004 07:48 Type: P Check: 7317
XX
1 CKKKKKKKKK KKKKKKKKK K
XX
!!AA SEQUENCE 1.0
ID ABG92659 standard; Protein; 58 AA.
XX
AC ABG92659;
XX
XX 18-NOV-2002 (first entry)
XX
XX Human DNA-binding protein #85.
XX
XX Human; DNA-binding protein; B cell immunodeficiency; autoimmune disorder;
XX severe combined immunodeficiency; rheumatoid arthritis; Crohn's disease;
XX diabetes mellitus; allergy; asthma; inflammatory condition; thrombosis;
XX graft-versus-host disease; blood-related disorder; atherosclerosis;
XX hyperproliferative disorder; cancer; renal disorder; arrhythmia;
XX acute glomerulonephritis; cardiovascular disorder; respiratory disorder;
XX Goodpasture's syndrome; neurological disorder; Alzheimer's disease;
XX Parkinson's disease; endocrine disorder; Addison's disease;
XX reproductive system disorder; endometriosis; infectious disease;
XX viral infection; bacterial infection; fungal infection; vaccine;
XX gastrointestinal disorder; multiple sclerosis; gene therapy.
XX
XX Homo sapiens.
XX
OS

```

PN US2002102638-A1.
XX 01-AUG-2002.
XX 17-JAN-2001; 2001US-0764846.
XX 31-JAN-2000; 2000US-179065P.
XX 04-FEB-2000; 2000US-180628P.
XX 28-JUN-2000; 2000US-214886P.
XX 07-JUL-2000; 2000US-216647P.
XX 07-JUL-2000; 2000US-216680P.
XX 11-JUL-2000; 2000US-217487P.
XX 11-JUL-2000; 2000US-217496P.
XX 14-JUL-2000; 2000US-218290P.
XX 26-JUL-2000; 2000US-220963P.
XX 26-JUL-2000; 2000US-220964P.
XX 14-AUG-2000; 2000US-224518P.
XX 14-AUG-2000; 2000US-224519P.
XX 14-AUG-2000; 2000US-225267P.
XX 14-AUG-2000; 2000US-225268P.
XX 14-AUG-2000; 2000US-225270P.
XX 14-AUG-2000; 2000US-225447P.
XX 14-AUG-2000; 2000US-225757P.
XX 14-AUG-2000; 2000US-225758P.
XX 22-AUG-2000; 2000US-226868P.
XX 30-AUG-2000; 2000US-228924P.
XX 01-SEP-2000; 2000US-229287P.
XX 01-SEP-2000; 2000US-229343P.
XX 01-SEP-2000; 2000US-229344P.
XX 01-SEP-2000; 2000US-229345P.
XX 05-SEP-2000; 2000US-229503P.
XX 08-SEP-2000; 2000US-229513P.
XX 08-SEP-2000; 2000US-231413P.
XX 21-SEP-2000; 2000US-234223P.
XX 21-SEP-2000; 2000US-234274P.
XX 25-SEP-2000; 2000US-234997P.
XX 27-SEP-2000; 2000US-235834P.
XX 29-SEP-2000; 2000US-236327P.
XX 29-SEP-2000; 2000US-236367P.
XX 29-SEP-2000; 2000US-236368P.
XX 29-SEP-2000; 2000US-236369P.
XX 29-SEP-2000; 2000US-236370P.
XX 02-OCT-2000; 2000US-236802P.
XX 02-OCT-2000; 2000US-237037P.
XX 02-OCT-2000; 2000US-237038P.
XX 02-OCT-2000; 2000US-237039P.
XX 02-OCT-2000; 2000US-237040P.
XX 13-OCT-2000; 2000US-239935P.
XX 20-OCT-2000; 2000US-240960P.
XX 20-OCT-2000; 2000US-241785P.
XX 20-OCT-2000; 2000US-241809P.
XX 01-NOV-2000; 2000US-244617P.
XX 17-NOV-2000; 2000US-249299P.
XX 08-DEC-2000; 2000US-251856P.
XX 08-DEC-2000; 2000US-251868P.
XX 08-DEC-2000; 2000US-251869P.
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
PI Rosen CA, Ruben SM, Barash SC;
XX WFI; 2002-690611/74.
XX N-PSDB; ABS68254.
XX Novel DNA-binding protein useful for diagnosis, prognosis, prevention
PT and treatment of immune, hyperproliferative, respiratory,
PT cardiovascular, reproductive, endocrine, gastrointestinal and
PT neurological disorders -
XX Claim 11; SEQ ID No 223; 225pp; English.
XX

CC The present invention relates to a new DNA-binding protein. The invention
CC is useful in treating, preventing, diagnosing and/or prognosing
CC immunodeficiencies (e.g. B cell immunodeficiencies, severe combined
CC immunodeficiencies), autoimmune disorders (rheumatoid arthritis, multiple
CC sclerosis, diabetes mellitus), allergic reactions and conditions (e.g.
CC asthma), inflammatory conditions, graft-versus-host disease, blood-
CC related disorders (thrombosis, atherosclerosis), hyperproliferative
CC disorders (e.g. cancer), renal disorders (e.g. acute glomerulonephritis),
CC cardiovascular disorders (e.g. arrhythmia), respiratory disorders
CC (Goodpasture's syndrome), neurological disorders (e.g. Alzheimer's
CC disease, Parkinson's disease), endocrine disorders (e.g. Addison's
CC disease), reproductive system disorders (e.g. endometriosis),
CC infectious diseases (e.g. viral, bacterial or fungal infections) and
CC gastrointestinal disorders (e.g. Crohn's disease). The invention is also
CC useful to stimulate neuronal growth and treat, prevent, and/or diagnose
CC neuronal damage which occurs in certain neuronal disorders or neuro-
CC degenerative conditions. The present amino acid sequence represents a
CC human DNA-binding protein of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from USPTO at
CC <http://seqdata.uspto.gov/sequence>.
XX
SQ Sequence 58 AA;
ABG92659 Length: 58 January 30, 2004 07:48 Type: P Check: 5509 ..
1 TYLECEHNSL VNSKCLTVVL SRCISVCLNK FYFVCKKKKK KKKKKKKKKK
51 KKKKKKKK
11AA SEQUENCE 1.0
ID AAU18238 standard; Protein; 58 AA.
XX
XX AAU18238;
XX 21-NOV-2001 (first entry)
XX Novel human DNA-binding protein #85.
XX Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; Y-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cytostatic.
XX
XX Homo sapiens.
XX
XX WO200155162-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01305.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.

```

!!SEQUENCE LIST 1.0
! FINDPATTERNS on geneseq:* allowing 0 mismatches
!      1 C(R,K){20,20}      January 30, 2004 07:38 ..

GENESEQP190S:AAW45801  ck: 9500  len: 39  finds: 1  ! Aaw45801 One chain of a bomb
GENESEQP200S:RAB13780  ck: 7317  len: 21  finds: 1  ! Aab13780 Soluble peptide and
GENESEQP200S:ABG92659  ck: 5509  len: 58  finds: 1  ! Abg92659 Human DNA-binding p
GENESEQP200S:AAU18238  ck: 5509  len: 58  finds: 1  ! Aau18238 Novel human DNA-bin
GENESEQP200S:AAO03766  ck: 8808  len: 81  finds: 1  ! Aao03766 Human polypeptide S
GENESEQP200S:AAO11210  ck: 863   len: 70  finds: 1  ! Aao11210 Human polypeptide S

```

\\End of list

Databases searched:
 Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds: 6
 Total length: 158,726,570
 Total sequences: 1,107,863
 CPU time: 07:02.46

```

! FINDPATTERNS on geneseq:* allowing 0 mismatches
!      1 C(R,K){20,20}      January 30, 2004 07:26 ...

AAW45801 ck: 9500 len: 39 ! Aaw45801 One chain of a bombesin dimer. 6/1
      C(R,K){20,20}
      C(K){20}
1:      CKKKKKKKKKKKKKKKKKKK XGGGQ

AAB13780 ck: 7317 len: 21 ! Aab13780 Soluble peptide antigen pK. 11/200
      C(R,K){20,20}
      C(K){20}
1:      CKKKKKKKKKKKKKKKKKKK

ABG92659 ck: 5509 len: 58 ! Abg92659 Human DNA-binding protein #85. 11/
      C(R,K){20,20}
      C(K){20}
35: KFYFV CKKKKKKKKKKKKKKKKKKK KKK

AAU18238 ck: 5509 len: 58 ! Aau18238 Novel human DNA-binding protein #8
      C(R,K){20,20}
      C(K){20}
35: KFYFV CKKKKKKKKKKKKKKKKKKK KKK

AAO03766 ck: 8808 len: 81 ! Aao03766 Human polypeptide SEQ ID NO 17658.
      C(R,K){20,20}
      C(K){20}
48: LTTTA CKKKKKKKKKKKKKKKKKKK KKKKK

AAO11210 ck: 863 len: 70 ! Aao11210 Human polypeptide SEQ ID NO 25102.
      C(R,K){20,20}
      C(K){20}
30: IDLCL CKKKKKKKKKKKKKKKKKKK KKKKK

```

Databases searched:
 Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds: 6
 Total length: 158,726,570
 Total sequences: 1,107,863
 CPU time: 04:40.14

SQ SEQUENCE 154 AA; 19596 MW; AEE8158A4839A216 CRC64;
Q8BHV2 Length: 154 January 30, 2004 08:18 Type: P Check: 8958 ..
1 MVQESRKQSN VGQKTSRRN VFKERVFRSR VPRDRRRGE EEEEEEEEE
51 EEEEEEEEE EEEEEEEEE KRERESEH EYSNANKDFG LLWIHPFCAC
101 DFMGRIFISH QRKNMAFLP SGDIIDRYLS YNWVMSLPS ILAYYMLKHC
151 GGCT

```

!!SEQUENCE LIST 1.0
! FINDPATTERNS on swp:* allowing 0 mismatches
!      1 (R,K){20,20}
January 30, 2004 07:59 ..
SP_FUN:Q12444 ck: 1384 len: 126 finds: 3 ! Q12444 saccharomyces cerevisia
SP_FUN:Q9P529 ck: 291 len: 128 finds: 13 ! Q9P529 neurospora crassa. hypc
SP_FUN:Q9NT34 ck: 7330 len: 380 finds: 7 ! Q9nt34 homo sapiens (human). h
SP_FUN:Q9H6Q7 ck: 3351 len: 720 finds: 10 ! Q9h6q7 homo sapiens (human). h
SP_FUN:Q8N6F0 ck: 9898 len: 55 finds: 15 ! Q8n6f0 homo sapiens (human). s
SP_FUN:Q9HC48 ck: 7602 len: 667 finds: 1 ! Q9hc48 homo sapiens (human). c
SP_FUN:Q9H5V6 ck: 379 len: 168 finds: 10 ! Q9h5v6 homo sapiens (human). h
SP_IN:Q8SWR7 ck: 2372 len: 515 finds: 4 ! Q8swr7 drosophila melanogaster
SP_IN:Q8T2U7 ck: 8768 len: 791 finds: 2 ! Q8t2u7 dictyostelium discoideu
SP_IN:Q8I247 ck: 5951 len: 206 finds: 2 ! Q8i247 plasmodium falciparum (c
SP_OM:Q95LV6 ck: 7515 len: 531 finds: 11 ! Q95lv6 macaca fascicularis (cr
SP_PL:Q9LXR2 ck: 4143 len: 517 finds: 4 ! Q9lxr2 arabidopsis thaliana (m
SP_PL:Q8S7D3 ck: 6479 len: 80 finds: 10 ! Q8s7d3 oryza sativa (rice). hy
SP_PL:Q8LQP6 ck: 3239 len: 113 finds: 9 ! Q8lqp6 oryza sativa (japonica
SP_PL:Q9LGZ9 ck: 6094 len: 260 finds: 222 ! Q9lgz9 arabidopsis thaliana (m
SP_RO:Q64075 ck: 8048 len: 215 finds: 3 ! Q64075 rattus sp. nucleoporin
SP_RO:Q9D5G1 ck: 9388 len: 169 finds: 6 ! Q9d5g1 mus musculus (mouse). a
SP_RO:Q35807 ck: 7510 len: 129 finds: 2 ! Q35807 rattus norvegicus (rat)
SP_RO:Q8BXG9 ck: 5434 len: 115 finds: 2 ! Q8bxg9 mus musculus (mouse). h
SP_RO:Q8BHV2 ck: 8958 len: 154 finds: 12 ! Q8bhv2 mus musculus (mouse). w
\\End of list

```

```

Databases searched:
SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

Total finds: 348
Total length: 305,079,309
Total sequences: 958,388
CPU time: 12:54.92

```

```

!!AA_SEQUENCE 1.0
ID_Q12444 PRELIMINARY; PRT; 126 AA.
AC Q12444;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF YOR309C.
OS YOR309C.
GN Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearson B.M., Hernando Y., Kalogeropoulos A., Schweizer M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=FY1679;
RA Pearson B.M., Hernando Y., Wolf S.S., Kalogeropoulos A., Schweizer M.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; 275217; CAA99629.1; -.
DR EMBL; X90565; CAA62164.1; -.
DR SGD; S0005836; YOR309C.
SQ SEQUENCE 126 AA; 16294 MW; 46EF1F4C664802C8 CRC64;

Q12444 Length: 126 January 30, 2004 08:18 Type: P Check: 1384 ..

1 NMQLPQRL LILNPILMMK RKRKRKRKR RERETMMKIP RILKLRKR
51 RTRRRKRKR RRRRRKRKR RRRRSRKR KRKNDAFYI LIISDPSRSL
101 LFGFRKFSII IQCLTYFSFH ILFHNH

!!AA_SEQUENCE 1.0
ID_Q9P529 PRELIMINARY; PRT; 128 AA.
AC Q9P529;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 15.2 kDa protein.
GN B24H17.160.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356815; CAB92638.2; -.
KW Hypothetical protein.
SQ SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFB70765 CRC64;

Q9P529 Length: 128 January 30, 2004 08:18 Type: P Check: 291 ..

1 MAIISGLHH KNINRRAPGH SVYSKNSSYD FQYDATQHY LPSQGFKAIP
51 DHLTGGKDC LSTHDKRKNQ KKKKKKKKK KKKKKKKKK KKKKKKKKK
101 KQEESRTYF QHFQADGIC PTPWHTR

!!AA_SEQUENCE 1.0
ID_Q9NT34 PRELIMINARY; PRT; 380 AA.

```

```

AC Q9NT34;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP43411120.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Otsenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137556; CAB70810.1; -.
DR Genew; HGNC:15736; C17orf28.
KW Hypothetical protein.
FT NON_TER 380 380
SQ SEQUENCE 380 AA; 42689 MW; 67F50DD101346AFB CRC64;

Q9NT34 Length: 380 January 30, 2004 08:18 Type: P Check: 7330 ..

1 MGSTDSKLNK RRAVIQLTTK TQPVETDDA FWDQFWADTA TSVQDVFAIV
51 PAEIRAVRE EPSNLTATLC YKAVEKLVOG AESGCHSEKE KQIVLNC SRL
101 LTRVLPIYFE DPMWRGFFWS TVPGAGRGGG EEDDEHARPL AESILLAIAD
151 LLFCPDFTVQ SHRSSTVDSA EDVHSLDSC EYWEAGVGFA HSPQPNVIHD
201 MNRMELLKLL LTCFSEAMYL PPAESGSGTN PWQFFCSTE NRHALPLFTS
251 LLNTVCAYDP VGYGIPYNHL LFSDYREPLV EEAQVLIVT LDHDSASSAS
301 PTVDGTTGT AMDADPPGP ENLFVNYLSR IHREEDFQFI LKGIARLLSN
351 LLLQKKKKKK KKKKKKKKK KKKKKKKKK

!!AA_SEQUENCE 1.0
ID_Q9H607 PRELIMINARY; PRT; 720 AA.
AC Q9H607;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ21979 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;
RL "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025632; BAB15196.1; -.
KW Hypothetical protein.
FT NON_TER 720 720
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Q9H607 Length: 720 January 30, 2004 08:18 Type: P Check: 3351 ..

1 MLTEQVEQYT KEWKNTCTII EDLKNELQKN KGASTLSQQT HMKIQSTLDI
51 LKEKTKAEER TAELEADAR EKDKELVEAL KRLKDYESGV YGLEDAVVEI
101 KNCKNQIKIR DREIEILTKE INKLEKISD FLDENEALRE RVGLEPKTMI
151 DLTEFRNSKH LKQOQYRAEN QILLKEISL EEEELDLKKK IRQMAQERGH
201 RSATSGLTTE DLNLTENISQ GDRISERKLD LLSLKNMSEA QSKNEFLSRE

```

SQ SEQUENCE 80 AA; 9362 MW; 0177C863133B21D8 CRC64;
Q8S7D3 Length: 80 January 30, 2004 08:18 Type: P Check: 6479 ..
1 MAVTMRQPVG RWSSEAAGGA TPGGRWGHWW WPAATAATRV LSVTHLDKCK
51 KKKKKKKKK KKKKKKKKK KKKKKKKKK

!!AA SEQUENCE 1.0
ID Q8LQ6 PRELIMINARY; PRT; 113 AA.
AC Q8LQ6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE OJ1117 G01.13 protein.
GN OJ1117 G01.13
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone:OJ1117.G01.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003374; BAB93330.1; -;
DR Gramene; O8LQ6; -;
SQ SEQUENCE 113 AA; 13660 MW; 597DB0EDB2BA3EF CRC64;

Q8LQ6 Length: 113 January 30, 2004 08:18 Type: P Check: 3239 ..
1 MEATSLSEHK KKKKKKKKK KKKKKKKKK REEBDEBEE
51 EEEELKNEY IWDIYEAAG INEKLMPGIV VYLAHVDECV RLRLTGVDVK
101 HEAWEICLTP VQV

!!AA SEQUENCE 1.0
ID Q9LG29 PRELIMINARY; PRT; 260 AA.
AC Q9LG29;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: FID9.
OS Arabidopsis thaliana (Mouset-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosidis II; Brassicales; Brassicaceae; Arabidopsids.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Nakamura Y.;
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002460; BAA97098.1; -;
DR InterPro; IPR005819; Histone_H5.
DR PRINTS; PR00624; HISTONEH5.
SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;

Q9LG29 Length: 260 January 30, 2004 08:18 Type: P Check: 6094 ..
1 MDCIRKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK
51 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK
101 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK
151 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK

201 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK
251 INWVGFVIF

!!AA SEQUENCE 1.0
ID Q64075 PRELIMINARY; PRT; 215 AA.
AC Q64075;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nucleoporin p62 homolog protein (Fragment).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95151924; PubMed=7849178;
RA Wang Z.Q., Akmal K.M., Kim K.H.;
RT "An unusual nucleoporin-related messenger ribonucleic acid is present
in the germ cells of rat testis.";
RL Biol. Reprod. 51:1022-1030(1994).
DR EMBL; S75997; AAB33384.1; -;
KW Porin.
FT NON TER 1 1
SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD88 CRC64;

Q64075 Length: 215 January 30, 2004 08:18 Type: P Check: 8048 ..
1 SGRATSSCD EDCSSSLPP SLSPGVKQDC EFLEKKKKK KKKKKKKKK
51 KKKKKKTGDN AKSVSRQYSL KVKLEHEAE QAKVELDFIL SQKELEDLL
101 SPLEESVKEQ SGTIYLQHAD EEREKTYKLA ENIDAQLKRM AQDLKDIIIEH
151 LNMAGGPADT SDPLQIQICKI LNAHMSLQW VDQSSALLQR RVEASRVCE
201 SRKQEQRSL RIAFD

!!AA SEQUENCE 1.0
ID Q9D5G1 PRELIMINARY; PRT; 169 AA.
AC Q9D5G1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Adult male testis cDNA, RIKEN full-length enriched library,
clone:4930444P10 product:hypothetical Arginine-rich region containing
protein, full insert sequence (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;

KW Hypothetical protein.
 SQ SEQUENCE 206 AA; 25047 MW; 1192E49A3DC4523F CRC64;
 Q81247 Length: 206 January 30, 2004 08:18 Type: P Check: 5951

1 MEGQHKEKNT KIKKSKPLV VSNRKPNNV EKKSAKPLV RDRPRDFSG
 51 SFNANFRNA YKPLYDSREQ EKKIIEKKLK SKNITBEKD ELKKYNDYK
 101 STDILLAKKE EERKLKAEVL KQEKQNLTK NKPYYYSR KIKKIYQEKL
 151 SSVKSLKKVI KKEKTKLQKE KGNIKPTKK KIFLKKKKKK KKKKKKKKK
 201 KKKKKT

!!AA SEQUENCE 1.0
 ID Q95LV6 PRELIMINARY; PRT; 531 AA.
 AC Q95LV6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical 61.4 kDa protein (Fragment).
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
 RA Terao K., Sugano S.;
 RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
 RT libraries."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB071085; BAB64479.1; -;
 KW Hypothetical protein.
 FT NON_TER 531 531
 SQ SEQUENCE 531 AA; 61389 MW; B55996B4F5CDD60C CRC64;
 Q95LV6 Length: 531 January 30, 2004 08:18 Type: P Check: 7515

1 MESESSNANM MNVQHEREDK NIQKMLPESV PCYSQHLSPS TYQMKDPPDC
 51 KSRSEPKSPE GRSSWNLSHI VQKTEQETHF RESVLEPISG YMMKQSPHMQ
 101 EGIKCVVGLK TSFPKTKGSE IGSMPHDPWP DENPRKWD SISEKTAWNP
 151 KNLQTVLKL DSSLMSSEY ESRVYTFEFL GKKSITSPKH VILKTKQLPI
 201 SQLFNIRCS TENHRKKQKH CFKYMKGQ WYTSIGBALR SATEYAKSPP
 251 SKSMIDKLLF NTAARGTSLN RTHQNVYGH TTEKEERVQE NVAASSLGPL
 301 DFPMPVLSDS KQNTWIRLS ERKTLNPKC LTMKEKKSPI SQIRKINRHP
 351 TTKHKKLES NLKTKLKAMW QGENVTDTPP NTISFTPDTS DIKRSRFTOT
 401 EIDWRISGLS HTQPTQIESL ABGIARCSDK RRTSNLVKGT KLHDSREGK
 451 KQHLTGMDP FAENPMTWT HLRKDPHLGK SEDVLLGEPP ISKSQYKGN
 501 SKKKKKKKKK KKKKKKKKKK KKKKKKKKK K

!!AA SEQUENCE 1.0
 ID Q9LXR2 PRELIMINARY; PRT; 517 AA.
 AC Q9LXR2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical 59.7 kDa protein.
 GN T20N10.250
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
 RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Arabidopsis sequencing project;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL353032; CAB88307.1; -;
 DR InterPro; IPR001810; F-box.
 DR InterPro; IPR006566; FBD.
 DR Pfam; PF00646; F-box; 1.
 DR SMART; SM00579; FBD; 1.
 DR SMART; SM00256; FBOX; 1.
 DR PROSITE; PS0181; FBOX; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 517 AA; 59689 MW; EC6D957D01F86E70 CRC64;
 Q9LXR2 Length: 517 January 30, 2004 08:18 Type: P Check: 4143

1 MDLFSLPNE LLYHLSPLS TKEAALTSVL SKRWNLPAF VPYLEFDDSV
 51 FLHPEERKE KEGILQSPMD FVDRVLDLHG DSLIKTFSLK CKTGVDSDHV
 101 DWICNVLAR GVSDDLDFID FRDLYSLPHE VGVSRLLVVL RVGESDLYW
 151 WQKFLCLPML KTLVLDSCWL CIGQFQILL ACPALEELDM TNTRWKDSNV
 201 TVSSSTLKEI TIDLHGCSV VNLKSLSFDA PSIVFYFYCD SLAEDYPQVN
 251 LKNLVEAQIN LLLTQAQIEQ VRALNEMLV ADDVFPGLN AKWLITGLRN
 301 VQQLYLPDT LEVLSRCEG MPVFNNLKV LSIWSDMNRGW QAMPVLLRNC
 351 PHELTLLIEG LLHYATDKCG DVCDICISRDY KDHSLTSCPV KKLQIYFRG
 401 TTRELEMIKH FLKIFPCLKE MDIYAHENSH TLFKDPITFE RVGKKKKKKK
 451 KKKKKKKKKK KKKKKKIRLN FRPVNKTEQF LKRLADKLCF IPQCLEFLDV
 501 DSSLGELALL AMDSRPS

!!AA SEQUENCE 1.0
 ID Q8S7D3 PRELIMINARY; PRT; 80 AA.
 AC Q8S7D3;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical 9.4 kDa protein.
 GN OSUNBA0057L21.23.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
 RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
 RA Tsitrin I., Riggs F., Heiao J., Zismann V., Blunt S., Pal G.,
 RA VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
 RA Salzberg S.L., White O., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBa0057L21 genomic sequence."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC087599; AAL79706.1; -;
 DR Gramene; Q8S7D3; -;
 KW Hypothetical protein.

```
DE GH22607p (Fragment).
GN CG7180.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN
RP
SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuno J., Pacieb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY095518; AAM12251.1; --
DR FlyBase; FBgn0032673; CG7180.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_PP.
DR Pfam; PF00102; Y_phosphatase; 1.
DR SMART; SM00194; PTPc; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00356; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.
FT NON TER 515 515
SQ SEQUENCE 515 AA; 59080 MW; B2825B7EBA96195E CRC64;

Q8SWR7 Length: 515 January 30, 2004 08:18 Type: P Check: 2372
1 MMTTQOLYVK PEAAENBEKS PAAASAAGA GADMTAATGG GSSGAGGGKT
51 GKHSRSSAR YDDVEKQKQK SRAIVSPNT IKLSMLNSGL LSPERIKLEA
101 RDENNSLSKT IPNGPIDIRH FLKCDLRK FVLYKLEFQ TAAKVESNTC
151 RHALKNNLE KNQPKCIPY DYNRVLEKV GGLQSDYVN ASYVDSLKP
201 NAYIVTQGPV EETVQAYWM VQENISAIV MLTKTFDPAK VMCHQYPPN
251 MEVHEQYQDI FINIVREQL ANFHIRTFL YKNEKQEVY DERLILQPHY
301 TEWYSHSCFP SNALLEFRR VRLVGNLIK DEBDMRGPII VHCSDDGGRS
351 GYVMSIDANL ELAEEBCFN VFGYKKLRQ SRKGLVENVE QYKFIYDTLE
401 EHIICKTWPF VSELSRLK AKARRNSGTK MNEYQAEYDQ ICKQTPRPTI
451 GDCAGGHRAD NREKNRDVLC VPPDNFRPYL TSFQGNFTD YIRKKKKKKK
501 KKKKKKKKKK KKKKK

!!AA_SEQUENCE 1.0
ID_Q8T2U7 PRELIMINARY; PRT; 791 AA.
AC
Q8T2U7
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Hypothetical 92.4 kDa protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN
RP
SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baugart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115574; AAL92183.1; --

DR InterPro; IPR005033; YEATS.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF03366; YEATS; 1.
DR SMART; SM00355; ZnF_C2H2; 1.
KW Hypothetical protein.
SQ SEQUENCE 791 AA; 92375 MW; D66CCB6DEC92352C CRC64;

Q8T2U7 Length: 791 January 30, 2004 08:18 Type: P Check: 8768
1 MKEKIEETKLL DTIDEKIENS TTTTITKNTH NNNNTNTNN NNNNNNNNN
51 NNNNNNNNN NNNNNNNNN NHPFSTLSAS LSKLRKLLII NNEGDRBELK
101 KRLITMIERE FPAELKNKEQ ELIEDDRIN TVKEMLSLN KORNSTKST
151 YHPIHTNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
201 NKKPRIFHK LQNGFVKLV CYCRRSDFV SGLGFLNHR IKGFYRSTL
251 DEAAEQIGVP VADSEIPKOD PSREGIVFRF KRGAYSPSIA FERNSEBDNG
301 TFDNNNNNN NNNNNNNNN NNNNNNTGN DTDKKNNG NDADVDMIDL
351 NVNYSNNK ETNEESSGSS RYVKKLIIV GNTSTQIHPD YRGHDSRTHK
401 WTVYVRGPQN EADISYFVK IWFIYLDSPA FNDKVEVVER PFNLTRRGWG
451 EFPVRIRLF HDKRNKPIDI IHNKLIQLP IQYVVPVVG ETTTEIDLR
501 LFPDKREKQ KLDQLNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
551 INNNNNNN SSNTSPTS YLNEPQKIVN NDKVSENSN NEDSQKNKEK
601 EKEKEKEKE EKEKEKEKE EKEKEKDK EKEKEKDK VKERDNKEKE
651 LENRPPDDR DRDREKERKE KKGKKGKDJE IEIEIETDIG IETEIGIETE
701 TETBIETGKE KEKEKEKEKE KEKEKEKEG KEIEIGRERG KKGKKEKEIE
751 TEMEIGKEIE ABIEIEKEK KKKKKKKKK KKKKKKKKK I

!!AA_SEQUENCE 1.0
ID_Q8I247 PRELIMINARY; PRT; 206 AA.
AC
Q8I247
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN PFA0475C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN
RP
SEQUENCE FROM N.A.
RX MEDLINE=2255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL031745; CAD49055.1; --
```

251 LIEKERDLER SRTVIKQFON KIKELVEBENK QLEEGMEKIL QAIKEMQKDP
 301 DVKGGETSLI IPSLERLVNA IESKNAEGIF DASLHLKAQV DQJTGNEEL
 351 ROELRESRKE AINYSQOLAK ANLKIDHLEK ETSLLRQSEG SNVVFKGIDL
 401 PDGIAPSSAS IINSONEYLI HLLQLEENKE KKLKNLEDSL EDYNRKFAVI
 451 RHQOSLYKE YLSEKETWKT ESKTIKEER KLEQVQODA IKVKEYNNLL
 501 NALQMSDEM KKLAESENK ITVLQVNEKS LIRQYTLIVE LERQLKENE
 551 KQKNELLSME ABCEKIGCL QRKEMAIK IALQKVVDN SVSLSELELA
 601 NKQVNEITAK YRDILQDNM LVQRTSNLEH LECENISLKE QVESINKLE
 651 ITREKLHTIE QAWBOETKLG NESSMDKAKK SITNSDIVSI SKKKKKKKK
 701 KKKKKKKKK KKKKKKKKK

!!AA SEQUENCE 1.0
 ID Q8N6F0 PRELIMINARY; PRT; 55 AA.

AC Q8N6F0;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Similar to LOC201361.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Lung;
 RA Strausberg R.;

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC030525; AH30525.1; -

SQ SEQUENCE 55 AA; 7251 MW; 0906032B284006BA CRC64;

Q8N6F0 Length: 55 January 30, 2004 08:18 Type: P Check: 9898 ..

1 MRLAQRRR GRGKKKKR KKKKKKKK KKKKKKKK KRRRRRGR

51 RRMQO

!!AA SEQUENCE 1.0

ID Q9HC48 PRELIMINARY; PRT; 667 AA.

AC Q9HC48;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CTCL tumor antigen se2-5 (Fragment).

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Testis;
 RX MEDLINE=21143360; PubMed=11149944;

RA Richmuller S., Usener D., Dummer R., Stein A., Thiel D.,
 RA Schendendorf D.;

RT "Serological detection of cutaneous T-cell lymphoma-associated
 antigens.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).

CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

DR EMBL; AF177228; AAG33676.1; -

DR HSSP; Q12923; 3PDZ.

DR InterPro; IPR001478; PDZ.

DR Pfam; PF00595; PDZ; 2.

DR SMART; SM00228; PDZ; 2.

DR PROSITE; PS0106; PDZ; 2.

FT NON_TER 1
 FT NON_TER 667
 SQ SEQUENCE 667 AA; 73499 MW; C653EC16802BAE02 CRC64;

Q9HC48 Length: 667 January 30, 2004 08:18 Type: P Check: 7602 ..

1 EHENLFREND CIVRINDGLD RNRFEQAOH MFRQAMRTPI IWFHVVPAA

51 KEQYEQLSQS EKNNYYSRF SPDSQYIDNR SVNSAGLTV QRAPRLNHP

101 EQIDSHSLRP HSAHPGKPP SAPASAPQNV FSTTVSSGN TKKIGKRLNI

151 QLKKGTEGLE FSITSRDVTI GGSAPIYVKN ILPRGAAIOD GRLKAGDRLI

201 EVNGVDLVGK SQEEVVSLLR STRMEGTSL LVFRQEDAFH PRELNAEPSQ

251 MQIPKETKAE DEDIVLTPDG TREPLTPEVP LNDGSGAGLG VSVKGNRSKE

301 NHADLGIFVK SIINGGAASK DGRLRVNDQL IAVNGESLIG KTNQDAMETL

351 RRSMTGKNGK RGMQLIVAR RISKCNELKS PGSPPGPELP IETALDDRER

401 RISHSLYSGI EGLDESPSRN AALSRIKES GKYQLSPTVN MPQDDTVILE

451 DDRLPVLPPH LSDQSSSSSH DDVGFTVADA GTWAKAAISD SADCSLSPDV

501 DPLAFQREG FCRQIADETK LNTVDDQKAG SPSRDVGPGL GLKSSSSLES

551 LQTAVAEVTL NGDIPFHRPR PRIIRGRGCN ESFRAAIDKS YDKPAVDDDD

601 EGMETLEEDT EESSRSGRES VSTASDQPSH SLERQNGNQ EKGDKTDRKK

651 KKKKKKKKK KKKKKKK

!!AA SEQUENCE 1.0

ID Q9HSV6 PRELIMINARY; PRT; 168 AA.

AC Q9HSV6;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Hypothetical protein FLJ22976 (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,

RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,

RA Nakamura Y., Isogai T., Sugano S.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK026629; BAB1513.1; -

KW Hypothetical protein.

FT NON_TER 168

SQ SEQUENCE 168 AA; 19549 MW; A19DBD195F8A1A90 CRC64;

Q9HSV6 Length: 168 January 30, 2004 08:18 Type: P Check: 379 ..

1 MNGGSRGLQ QKGNVDGVA ATPTAASASC QYRCIECQOE AKELYRDVNH

51 GVLKTIKCS CQKPVDKYIE YDPVILINA ILCKAQAYRH ILFNTQINIH

101 GKLCIFCLLC EAYLRWMLQ DSNQNTAPDD LIRYVREWEK KKKKKKKKK

151 KKKKKKKKK KKKKKKK

!!AA SEQUENCE 1.0

ID Q8SWR7 PRELIMINARY; PRT; 515 AA.

AC Q8SWR7;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

196: KKKKK (K){20} KKKKK
 197: KKKKK (K){20} KKKKK
 198: KKKKK (K){20} KKKKK
 199: KKKKK (K){20} KKKKK
 200: KKKKK (K){20} KKKKK
 201: KKKKK (K){20} KKKKK
 202: KKKKK (K){20} KKKKK
 203: KKKKK (K){20} KKKKK
 204: KKKKK (K){20} KKKKK
 205: KKKKK (K){20} KKKKK
 206: KKKKK (K){20} KKKKK
 207: KKKKK (K){20} KKKKK
 208: KKKKK (K){20} KKKKK
 209: KKKKK (K){20} KKKKK
 210: KKKKK (K){20} KKKKK
 211: KKKKK (K){20} KKKKK
 212: KKKKK (K){20} KKKKK
 213: KKKKK (K){20} KKKKK
 214: KKKKK (K){20} KKKKK
 215: KKKKK (K){20} KKKKK
 216: KKKKK (K){20} KKKKK
 217: KKKKK (K){20} KKKKK
 218: KKKKK (K){20} KKKKK
 219: KKKKK (K){20} KKKKK

220: KKKKK (K){20} KKKKK
 221: KKKKK (K){20} KKKKK
 222: KKKKK (K){20} KKKKK
 223: KKKKK (K){20} KKKKK
 224: KKKKK (K){20} KKKKK
 225: KKKKK (K){20} KKKKK
 226: KKKKK (K){20} KKKKK
 227: KKKKK (K){20} KKKKK
 228: KKKKK (K){20} KKKKK
 229: KKKKK (K){20} KKKKK
 230: KKKKK (K){20} KKKKK
 231: KKKKK (K){20} KKKKK
 232: KKKKK (K){20} KKKKK
 233: KKKKK (K){20} KKKKK
 234: KKKKK (K){20} KKKKK
 235: KKKKK (K){20} KKKKK
 236: KKKKK (K){20} KKKKK
 237: KKKKK (K){20} KKKKK
 238: KKKKK (K){20} KKKKK
 239: KKKKK (K){20} KKKKK
 240: KKKKK (K){20} KKKKK
 241: KKKKK (K){20} KKKKK
 242: KKKKK (K){20} KKKKK
 243: KKKKK (K){20} KKKKK

293: KKKKK (K) {20} KKKKK
294: KKKKK (K) {20} KKKKK
295: KKKKK (K) {20} KKKKK
296: KKKKK (K) {20} KKKKK
297: KKKKK (K) {20} KKKKK
298: KKKKK (K) {20} KKKKK
299: KKKKK (K) {20} KKKKK
300: KKKKK (K) {20} KKKKK
301: KKKKK (K) {20} KKKKK
302: KKKKK (K) {20} KKKKK
303: KKKKK (K) {20} KKKKK
304: KKKKK (K) {20} KKKKK
305: KKKKK (K) {20} KKKKK
306: KKKKK (K) {20} KKKKK
307: KKKKK (K) {20} KKKKK
308: KKKKK (K) {20} KKKKK
309: KKKKK (K) {20} KKKKK
310: KKKKK (K) {20} KKKKK
311: KKKKK (K) {20} KKKKK
312: KKKKK (K) {20} KKKKK
313: KKKKK (K) {20} KKKKK
314: KKKKK (K) {20} KKKKK
315: KKKKK (K) {20} KKKKK
316: KKKKK (K) {20} KKKKK

317: KKKKK (K) {20} KKKKK
318: KKKKK (K) {20} KKKKK
319: KKKKK (K) {20} KKKKK
320: KKKKK (K) {20} KKKKK
321: KKKKK (K) {20} KKKKK
322: KKKKK (K) {20} KKKKK
323: KKKKK (K) {20} KKKKK
324: KKKKK (K) {20} KKKKK
325: KKKKK (K) {20} KKKKK
326: KKKKK (K) {20} KKKKK
327: KKKKK (K) {20} KKKKK
328: KKKKK (K) {20} KKKKK
329: KKKKK (K) {20} KKKKK
330: KKKKK (K) {20} KKKKK
331: KKKKK (K) {20} KKKKK
332: KKKKK (K) {20} KKKKK
333: KKKKK (K) {20} KKKKK
334: KKKKK (K) {20} KKKKK
335: KKKKK (K) {20} KKKKK
336: KKKKK (K) {20} KKKKK
337: KKKKK (K) {20} KKKKK
338: KKKKK (K) {20} KKKKK
339: KKKKK (K) {20} KKKKK
340: KKKKK (K) {20} KKKKK
341: KKKKK (K) {20} KKKKK

342: KKKKK (K){20} KKKKK
343: KKKKK (K){20} KKKKK
344: KKKKK (K){20} KKKKK
345: KKKKK (K){20} KKKKK
346: KKKKK (K){20} KKKKK
347: KKKKK (K){20} KKKKK
348: KKKKK (K){20} KKKKK
349: KKKKK (K){20} KKKKK
350: KKKKK (K){20} KKKKK
351: KKKKK (K){20} KKKKK
352: KKKKK (K){20} KKKKK
353: KKKKK (K){20} KKKKK
354: KKKKK (K){20} KKKKK
355: KKKKK (K){20} KKKKK
356: KKKKK (K){20} KKKKK
357: KKKKK (K){20} KKKKK
358: KKKKK (K){20} KKKKK
359: KKKKK (K){20} KKKKK
360: KKKKK (K){20} KKKKK
361: KKKKK (K){20} KKKKK
362: KKKKK (K){20} KKKKK
363: KKKKK (K){20} KKKKK
364: KKKKK (K){20} KKKKK
365: KKKKK (K){20} KKKKK

366: KKKKK (K){20} KKKKK
367: KKKKK (K){20} KKKKK
368: KKKKK (K){20} KKKKK
369: KKKKK (K){20} KKKKK
370: KKKKK (K){20} KKKKK
371: KKKKK (K){20} KKKKK
372: KKKKK (K){20} KKKKK
373: KKKKK (K){20} KKKKK
374: KKKKK (K){20} KKKKK
375: KKKKK (K){20} KKKKK
376: KKKKK (K){20} KKKKK
377: KKKKK (K){20} KKKKK
378: KKKKK (K){20} KKKKK
379: KKKKK (K){20} KKKKK
380: KKKKK (K){20} KKKKK
381: KKKKK (K){20} KKKKK
382: KKKKK (K){20} KKKKK
383: KKKKK (K){20} KKKKK
384: KKKKK (K){20} KKKKK
385: KKKKK (K){20} KKKKK
386: KKKKK (K){20} KKKKK
387: KKKKK (K){20} KKKKK
388: KKKKK (K){20} KKKKK
389: KKKKK (K){20} KKKKK

37: REKK KKKKKKKKKKKKKKKKKKKKK
AAW45801 ck: 9500 len: 39 ! Aaw45801 One chain of a bombesin dimer. 6/1
(R,K){20,20}
(K){20}
2: C KKKKKKKKKKKKKKKKKKKKK XGGGQ
AAW45800 ck: 9056 len: 39 ! Aaw45800 One chain of an alpha-melanocyte s
(R,K){20,20}
(K){20}
18: VGGGX KKKKKKKKKKKKKKKKKKKKK CK
AAW45802 ck: 7664 len: 35 ! Aaw45802 One chain of an alpha-MSH receptor
(R,K){20,20}
(K){20}
14: VGGGX KKKKKKKKKKKKKKKKKKKKK CK
AAW43246 ck: 9752 len: 32 ! Aay43246 Cell-surface molecule binding pep
(R,K){20,20}
(K){20}
12: SSGGS KKKKKKKKKKKKKKKKKKKKK K
(K){20}
13: GSGSK KKKKKKKKKKKKKKKKKKKKK
AAW07213 ck: 2211 len: 40 ! Aay07213 Peptide transfection vector #1. 7/
(R,K){20,20}
(K){20}
21: YEDES KKKKKKKKKKKKKKKKKKKKK
AAW12950 ck: 5821 len: 62 ! Aay12950 Amino acid sequence of a human seq
(R,K){20,20}
(K){20}
40: QFQAS KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
41: FQASK KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
42: QASKK KKKKKKKKKKKKKKKKKKKKK K
(K){20}
43: ASKKK KKKKKKKKKKKKKKKKKKKKK
AAB59105 ck: 8456 len: 27 ! Aab59105 Breast and ovarian cancer associat
(R,K){20,20}
(K){20}
6: NSAXX KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
7: SAXXK KKKKKKKKKKKKKKKKKKKKK K
(K){20}
8: AXXKK KKKKKKKKKKKKKKKKKKKKK
AAB53249 ck: 4945 len: 59 ! Aab53249 Human colon cancer antigen protein
(R,K){20,20}

27: ALLSL KKKKKKKKKKKKKKKKKKKKK (K){20}
28: LLSLK KKKKKKKKKKKKKKKKKKKKK (K){20}
29: LSLKK KKKKKKKKKKKKKKKKKKKKK (K){20}
30: SLKKK KKKKKKKKKKKKKKKKKKKKK (K){20}
31: LKKKK KKKKKKKKKKKKKKKKKKKKK (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20}
AAB53659 ck: 3850 len: 184 ! Aab53659 Human colon cancer antigen protei
(R,K){20,20}
(K){20}
1: KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
AAB53800 ck: 296 len: 69 ! Aab53800 Human colon cancer antigen protein
(R,K){20,20}
(K){20}
20: NVDCS KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: VDCEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: DCEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
23: CEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

9: CSFQK KKKKKKKKKKKKKKKKKKKKK KKKKG
(K) {20}
10: SFQKK KKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
11: FQKKK KKKKKKKKKKKKKKKKKKKKK KKKGR
(K) {20}
12: QKKKK KKKKKKKKKKKKKKKKKKKKK XGSRP
(K) {20}
AAB54314 ck: 6235 len: 55 ! Aab54314 Human pancreatic cancer antigen pr
(R,K) {20,20}
24: FFXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
25: FXXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
26: XXXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
27: XXKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
28: XXXKK KKKKKKKKKKKKKKKKKKKKK KKKKG
(K) {20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGR
(K) {20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGRP
(K) {20}
32: KKKKK KKKKKKKKKKKKKKKKKKKKK GGRF
(K) {20}

AAB56121 ck: 5941 len: 125 ! Aab56121 Human secreted protein sequence en
(R,K) {20,20}
87: ILKSQ KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
88: LKSQK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
89: KSQKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
90: SQKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
91: QKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
92: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
93: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
94: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
95: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

1

1

96: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
97: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
98: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
99: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKG
(K) {20}
100: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
101: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGP
(K) {20}
102: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGP
(K) {20}
103: KKKKK KKKKKKKKKKKKKKKKKKKKK GGP
(K) {20}

AAB27956 ck: 6732 len: 139 ! Aab27956 Human secreted protein SEQ ID NO: 1
(R,K) {20,20}
91: HFRPG KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
92: FRPGK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
93: RPKGK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
94: PGKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
95: GKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
96: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
97: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
98: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
99: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
100: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
101: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
102: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
103: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
104: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
105: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

106: KKKKK (K){20} KKKKKK
107: KKKKK (K){20} KKKKKK
108: KKKKK (K){20} KKKKKK
109: KKKKK (K){20} KKKKKK
110: KKKKK (K){20} KKKKKK
111: KKKKK (K){20} KKKKKK
112: KKKKK (K){20} KKKKKK
113: KKKKK (K){20} KKKKKK
114: KKKKK (K){20} KKKKKK
115: KKKKK (K){20} KKKKKK
116: KKKKK (K){20} KKKKKK
117: KKKKK (K){20} KKKKKK
118: KKKKK (K){20} KKKKKK
119: KKKKK (K){20} KKKKKK
120: KKKKK (K){20} KKKKKK
AAB39140 ck: 9531 len: 66 | Aab39140 Human secreted protein #48. 2/2001
41: MWTVX KKKKK (R,K){20,20} KKKKK
42: WTVXK KKKKK (K){20} KKKKK
43: TVXKK KKKKK (K){20} KKKKK
44: VXKKK KKKKK (K){20} KKKKK
45: XKKKK KKKKK (K){20} KKKKK
46: KKKKK KKKKK (K){20} KKKKK
47: KKKKK KKKKK (K){20} KKKKK

AAB41457 ck: 4536 len: 168 | Aab41457 Human ORFX ORF1221 polypeptide seq

1
144: KQKPK (R,K){20,20} KKKKKK
145: QKPKK (K){20} KKKKKK
146: QKPKK (K){20} KKKKKK
147: KPKKK (K){20} KKKKKK
148: PKKKK (K){20} KKKKKK
149: KKKKK (K){20} KKKKKK

AAB42786 ck: 4959 len: 102 | Aab42786 Human ORFX ORF2550 polypeptide se

1
81: SDVLQ KKKKK (R,K){20,20} KKKKKK
82: DVLQK KKKKK (K){20} KKKKKK
83: VLQKK KKKKK (K){20} KKKKKK

AAB42889 ck: 6432 len: 62 | Aab42889 Human ORFX ORF2653 polypeptide se

1
40: SHRCL KKKKK (R,K){20,20} KKKKKK
41: HRCLK KKKKK (K){20} KKKKKK
42: RCLKK KKKKK (K){20} KKKKKK
43: CLKKK KKKKK (K){20} KKKKKK

AAB43641 ck: 5216 len: 133 | Aab43641 Human cancer associated protein s

1
114: GHEQS KKKKK (R,K){20,20} KKKKKK

AAB43835 ck: 4025 len: 223 | Aab43835 Human cancer associated protein s

1
196: NILFW KKKKK (R,K){20,20} KKKKKK
 KXGGA

AAB44188 ck: 4991 len: 43 | Aab44188 Human cancer associated protein s

1
2: E KKKKK (R,K){20,20} KKKKKK
3: EK KKKKK (K){20} KKKKKK

4: EKK KKKKKKKKKKKKKKKKKKKKK KKKGR
(K) {20}
5: EKKK KKKKKKKKKKKKKKKKKKKKK KGRPP
(K) {20}
6: EKKKK KKKKKKKKKKKKKKKKKKKKK KGRPP
(K) {20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKK GRPEX
(K) {20}

AAB44380 ck: 8490 len: 42 ! Aab44380 Human secreted protein encoded by
(R,K) {20,20}
(K) {20}

22: IWKKI KKKKKKKKKKKKKKKKKKKKK K
(K) {20}

23: WKTK KKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAB34331 ck: 8131 len: 66 ! Aab34331 Human secreted protein sequence en
(R,K) {20,20}
(K) {20}

40: IHHPK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

41: HHHPK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

42: HKPKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

43: KPKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

44: PKKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K) {20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKK KK
(K) {20}

46: KKKKK KKKKKKKKKKKKKKKKKKKKK K
(K) {20}

47: KKKKK KKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAB23585 ck: 9893 len: 36 ! Aab23585 Ask21 linker peptide. 1/2001
(R,K) {20,20}
(K) {20}

15: PGSGS KKKKKKKKKKKKKKKKKKKKK KG
(K) {20}

16: GSGSK KKKKKKKKKKKKKKKKKKKKK G
(K) {20}

AAB23586 ck: 58 len: 36 ! Aab23586 Gsk21 linker peptide. 1/2001
(R,K) {20,20}
(K) {20}

15: SGSGS KKKKKKKKKKKKKKKKKKKKK KG
(K) {20}

16: GSGSK KKKKKKKKKKKKKKKKKKKKK G
(K) {20}

AAB23591 ck: 7684 len: 630 ! Aab23591 Modified fibre protein encoded in
(R,K) {20,20}

596: PGSGS KKKKKKKKKKKKKKKKKKKKK KGSYS
(K) {20}
597: GSGSK KKKKKKKKKKKKKKKKKKKKK GSYSM
(K) {20}

AAB23592 ck: 7647 len: 630 ! Aab23592 Modified fibre protein encoded in
(R,K) {20,20}
(K) {20}

596: SGSGS KKKKKKKKKKKKKKKKKKKKK KGSYS
(K) {20}

597: GSGSK KKKKKKKKKKKKKKKKKKKKK GSYSM
(K) {20}

AAB23593 ck: 2249 len: 640 ! Aab23593 Modified fibre protein encoded in
(R,K) {20,20}
(K) {20}

596: PGSGS KKKKKKKKKKKKKKKKKKKKK KGSAB
(K) {20}

597: GSGSK KKKKKKKKKKKKKKKKKKKKK GSABK
(K) {20}

AAB23594 ck: 2612 len: 640 ! Aab23594 Modified fibre protein encoded in
(R,K) {20,20}
(K) {20}

596: SGSGS KKKKKKKKKKKKKKKKKKKKK KGSAB
(K) {20}

597: GSGSK KKKKKKKKKKKKKKKKKKKKK GSABK
(K) {20}

AAB13780 ck: 7317 len: 21 ! Aab13780 Soluble peptide antigen pK. 11/2001
(R,K) {20,20}
(K) {20}

2: C KKKKKKKKKKKKKKKKKKKKK

AAB13783 ck: 4553 len: 45 ! Aab13783 Soluble tandem pEA/ pK peptide cor
(R,K) {20,20}
(K) {20}

26: AAAAA KKKKKKKKKKKKKKKKKKKKK

AAB13784 ck: 4126 len: 44 ! Aab13784 Soluble tandem HA/ pK peptide con
(R,K) {20,20}
(K) {20}

25: DGYWG KKKKKKKKKKKKKKKKKKKKK

AAG00834 ck: 6330 len: 103 ! Aag00834 Human secreted protein, SEQ ID NO:
(R,K) {20,20}
(K) {20}

83: RKRQQ KKKKKKKKKKKKKKKKKKKKK K
(K) {20}

84: KRQOK KKKKKKKKKKKKKKKKKKKKK
(K) {20}

AAY98493 ck: 8137 len: 45 ! Aay98493 Peptide #5 used in nucleic acid tr
(R,K) {20,20}
(K) {20}

4: YKA KKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K) {20} KKKKK
 7: KKKKK (K) {20} KKKKK
 8: KKKKK (K) {20} KKKKK
 9: KKKKK (K) {20} KKKKK
 10: KKKKK (K) {20} KKKKK
 11: KKKKK (K) {20} KKKKK
 12: KKKKK (K) {20} KKKKK
 13: KKKKK (K) {20} KKKKK
 14: KKKKK (K) {20} KKKKK
 15: KKKKK (K) {20} KKKKK
 16: KKKKK (K) {20} KKKKK
 17: KKKKK (K) {20} KKKKK
 18: KKKKK (K) {20} KKKKK
 19: KKKKK (K) {20} KKKKK
 20: KKKKK (K) {20} KKKKK
 21: KKKKK (K) {20} KKKKK
 22: KKKKK (K) {20} KKKKK
 23: KKKKK (K) {20} KKKKK
 24: KKKKK (K) {20} KKKKK
 25: KKKKK (K) {20} KKKKK
 26: KKKKK (K) {20} KKKKK
 27: KKKKK (K) {20} KKKKK
 28: KKKKK (K) {20} KKKKK
 29: KKKKK (K) {20} KKKKK

30: KKKKK (K) {20} KKKKK
 31: KKKKK (K) {20} KKKKK
 32: KKKKK (K) {20} KKKKK
 33: KKKKK (K) {20} KKKKK
 34: KKKKK (K) {20} KKKKK
 35: KKKKK (K) {20} KKKKK
 36: KKKKK (K) {20} KKKKK
 37: KKKKK (K) {20} KKKKK
 38: KKKKK (K) {20} KKKKK
 39: KKKKK (K) {20} KKKKK
 40: KKKKK (K) {20} KKKKK
 41: KKKKK (K) {20} KKKKK
 42: KKKKK (K) {20} KKKKK
 43: KKKKK (K) {20} KKKKK
 44: KKKKK (K) {20} KKKKK
 45: KKKKK (K) {20} KKKKK
 46: KKKKK (K) {20} KKKKK
 47: KKKKK (K) {20} KKKKK
 48: KKKKK (K) {20} KKKKK
 49: KKKKK (K) {20} KKKKK
 50: KKKKK (K) {20} KKKKK
 51: KKKKK (K) {20} KKKKK
 52: KKKKK (K) {20} KKKKK
 53: KKKKK (K) {20} KKKKK

23: KKKKK (K){20} KKKKK
24: KKKKK (K){20} KKKGG
25: KKKKK (K){20} KKGGR
26: KKKKK (K){20} KGGRR
27: KKKKK (K){20} GGRRG

AAy56902 ck: 4875 len: 30 ! Aay56902 (Lys)30 protein sequence. 4/2000

1

1: (R,K){20,20} KKKKK
2: K KKKKK (K){20} KKKKK
3: KK KKKKK (K){20} KKKKK
4: KKK KKKKK (K){20} KKKKK
5: KKKK KKKKK (K){20} KKKKK
6: KKKKK KKKKK (K){20} KKKKK
7: KKKKK KKKKK (K){20} KKKK
8: KKKKK KKKKK (K){20} KKK
9: KKKKK KKKKK (K){20} KK
10: KKKKK KKKKK (K){20} K
11: KKKKK KKKKK (K){20}

AAy56903 ck: 5075 len: 434 ! Aay56903 (Lys)434 protein sequence. 4/2000

1

1: (R,K){20,20} KKKKK
2: K KKKKK (K){20} KKKKK
3: KK KKKKK (K){20} KKKKK
4: KKK KKKKK (K){20} KKKKK
5: KKKK KKKKK (K){20} KKKKK

6: KKKKK KKKKK (K){20} KKKKK
7: KKKKK KKKKK (K){20} KKKKK
8: KKKKK KKKKK (K){20} KKKKK
9: KKKKK KKKKK (K){20} KKKKK
10: KKKKK KKKKK (K){20} KKKKK
11: KKKKK KKKKK (K){20} KKKKK
12: KKKKK KKKKK (K){20} KKKKK
13: KKKKK KKKKK (K){20} KKKKK
14: KKKKK KKKKK (K){20} KKKKK
15: KKKKK KKKKK (K){20} KKKKK
16: KKKKK KKKKK (K){20} KKKKK
17: KKKKK KKKKK (K){20} KKKKK
18: KKKKK KKKKK (K){20} KKKKK
19: KKKKK KKKKK (K){20} KKKKK
20: KKKKK KKKKK (K){20} KKKKK
21: KKKKK KKKKK (K){20} KKKKK
22: KKKKK KKKKK (K){20} KKKKK
23: KKKKK KKKKK (K){20} KKKKK
24: KKKKK KKKKK (K){20} KKKKK
25: KKKKK KKKKK (K){20} KKKKK
26: KKKKK KKKKK (K){20} KKKKK
27: KKKKK KKKKK (K){20} KKKKK
28: KKKKK KKKKK (K){20} KKKKK
29: KKKKK KKKKK (K){20} KKKKK
30: KKKKK KKKKK (K){20} KKKKK

31: (K) {20} XXXXX
32: (K) {20} XXXXX
33: (K) {20} XXXXX
34: (K) {20} XXXXX
35: (K) {20} XXXXX
36: (K) {20} XXXXX
37: (K) {20} XXXXX
38: (K) {20} XXXXX
39: (K) {20} XXXXX
40: (K) {20} XXXXX
41: (K) {20} XXXXX
42: (K) {20} XXXXX
43: (K) {20} XXXXX
44: (K) {20} XXXXX
45: (K) {20} XXXXX
46: (K) {20} XXXXX
47: (K) {20} XXXXX
48: (K) {20} XXXXX
49: (K) {20} XXXXX
50: (K) {20} XXXXX
51: (K) {20} XXXXX
52: (K) {20} XXXXX
53: (K) {20} XXXXX
54: (K) {20} XXXXX

55: (K) {20} XXXXX
56: (K) {20} XXXXX
57: (K) {20} XXXXX
58: (K) {20} XXXXX
59: (K) {20} XXXXX
60: (K) {20} XXXXX
61: (K) {20} XXXXX
62: (K) {20} XXXXX
63: (K) {20} XXXXX
64: (K) {20} XXXXX
65: (K) {20} XXXXX
66: (K) {20} XXXXX
67: (K) {20} XXXXX
68: (K) {20} XXXXX
69: (K) {20} XXXXX
70: (K) {20} XXXXX
71: (K) {20} XXXXX
72: (K) {20} XXXXX
73: (K) {20} XXXXX
74: (K) {20} XXXXX
75: (K) {20} XXXXX
76: (K) {20} XXXXX
77: (K) {20} XXXXX
78: (K) {20} XXXXX

128: KKKKK (K) {20} KKKKK
129: KKKKK (K) {20} KKKKK
130: KKKKK (K) {20} KKKKK
131: KKKKK (K) {20} KKKKK
132: KKKKK (K) {20} KKKKK
133: KKKKK (K) {20} KKKKK
134: KKKKK (K) {20} KKKKK
135: KKKKK (K) {20} KKKKK
136: KKKKK (K) {20} KKKKK
137: KKKKK (K) {20} KKKKK
138: KKKKK (K) {20} KKKKK
139: KKKKK (K) {20} KKKKK
140: KKKKK (K) {20} KKKKK
141: KKKKK (K) {20} KKKKK
142: KKKKK (K) {20} KKKKK
143: KKKKK (K) {20} KKKKK
144: KKKKK (K) {20} KKKKK
145: KKKKK (K) {20} KKKKK
146: KKKKK (K) {20} KKKKK
147: KKKKK (K) {20} KKKKK
148: KKKKK (K) {20} KKKKK
149: KKKKK (K) {20} KKKKK
150: KKKKK (K) {20} KKKKK
151: KKKKK (K) {20} KKKKK

152: KKKKK (K) {20} KKKKK
153: KKKKK (K) {20} KKKKK
154: KKKKK (K) {20} KKKKK
155: KKKKK (K) {20} KKKKK
156: KKKKK (K) {20} KKKKK
157: KKKKK (K) {20} KKKKK
158: KKKKK (K) {20} KKKKK
159: KKKKK (K) {20} KKKKK
160: KKKKK (K) {20} KKKKK
161: KKKKK (K) {20} KKKKK
162: KKKKK (K) {20} KKKKK
163: KKKKK (K) {20} KKKKK
164: KKKKK (K) {20} KKKKK
165: KKKKK (K) {20} KKKKK
166: KKKKK (K) {20} KKKKK
167: KKKKK (K) {20} KKKKK
168: KKKKK (K) {20} KKKKK
169: KKKKK (K) {20} KKKKK
170: KKKKK (K) {20} KKKKK
171: KKKKK (K) {20} KKKKK
172: KKKKK (K) {20} KKKKK
173: KKKKK (K) {20} KKKKK
174: KKKKK (K) {20} KKKKK
175: KKKKK (K) {20} KKKKK
176: KKKKK (K) {20} KKKKK

177: KKKKK (K) {20} KKKKK
178: KKKKK (K) {20} KKKKK
179: KKKKK (K) {20} KKKKK
180: KKKKK (K) {20} KKKKK
181: KKKKK (K) {20} KKKKK
182: KKKKK (K) {20} KKKKK
183: KKKKK (K) {20} KKKKK
184: KKKKK (K) {20} KKKKK
185: KKKKK (K) {20} KKKKK
186: KKKKK (K) {20} KKKKK
187: KKKKK (K) {20} KKKKK
188: KKKKK (K) {20} KKKKK
189: KKKKK (K) {20} KKKKK
190: KKKKK (K) {20} KKKKK
191: KKKKK (K) {20} KKKKK
192: KKKKK (K) {20} KKKKK
193: KKKKK (K) {20} KKKKK
194: KKKKK (K) {20} KKKKK
195: KKKKK (K) {20} KKKKK
196: KKKKK (K) {20} KKKKK
197: KKKKK (K) {20} KKKKK
198: KKKKK (K) {20} KKKKK
199: KKKKK (K) {20} KKKKK
200: KKKKK (K) {20} KKKKK

201: KKKKK (K) {20} KKKKK
202: KKKKK (K) {20} KKKKK
203: KKKKK (K) {20} KKKKK
204: KKKKK (K) {20} KKKKK
205: KKKKK (K) {20} KKKKK
206: KKKKK (K) {20} KKKKK
207: KKKKK (K) {20} KKKKK
208: KKKKK (K) {20} KKKKK
209: KKKKK (K) {20} KKKKK
210: KKKKK (K) {20} KKKKK
211: KKKKK (K) {20} KKKKK
212: KKKKK (K) {20} KKKKK
213: KKKKK (K) {20} KKKKK
214: KKKKK (K) {20} KKKKK
215: KKKKK (K) {20} KKKKK
216: KKKKK (K) {20} KKKKK
217: KKKKK (K) {20} KKKKK
218: KKKKK (K) {20} KKKKK
219: KKKKK (K) {20} KKKKK
220: KKKKK (K) {20} KKKKK
221: KKKKK (K) {20} KKKKK
222: KKKKK (K) {20} KKKKK
223: KKKKK (K) {20} KKKKK
224: KKKKK (K) {20} KKKKK

298:	XXXX	XXXXXXXXXXXXXXXXXXXX	XXXX	XXXX
299:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
300:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
301:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
302:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
303:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
304:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
305:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
306:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
307:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
308:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
309:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
310:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
311:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
312:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
313:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
314:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
315:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
316:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
317:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
318:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
319:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
320:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
321:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX
322:	XXXX	(K) {20}	XXXXXXXXXXXX	XXXX

323: KKKKK (K) {20} KKKKK
 324: KKKKK (K) {20} KKKKK
 325: KKKKK (K) {20} KKKKK
 326: KKKKK (K) {20} KKKKK
 327: KKKKK (K) {20} KKKKK
 328: KKKKK (K) {20} KKKKK
 329: KKKKK (K) {20} KKKKK
 330: KKKKK (K) {20} KKKKK
 331: KKKKK (K) {20} KKKKK
 332: KKKKK (K) {20} KKKKK
 333: KKKKK (K) {20} KKKKK
 334: KKKKK (K) {20} KKKKK
 335: KKKKK (K) {20} KKKKK
 336: KKKKK (K) {20} KKKKK
 337: KKKKK (K) {20} KKKKK
 338: KKKKK (K) {20} KKKKK
 339: KKKKK (K) {20} KKKKK
 340: KKKKK (K) {20} KKKKK
 341: KKKKK (K) {20} KKKKK
 342: KKKKK (K) {20} KKKKK
 343: KKKKK (K) {20} KKKKK
 344: KKKKK (K) {20} KKKKK
 345: KKKKK (K) {20} KKKKK
 346: KKKKK (K) {20} KKKKK

347: KKKKK (K) {20} KKKKK
 348: KKKKK (K) {20} KKKKK
 349: KKKKK (K) {20} KKKKK
 350: KKKKK (K) {20} KKKKK
 351: KKKKK (K) {20} KKKKK
 352: KKKKK (K) {20} KKKKK
 353: KKKKK (K) {20} KKKKK
 354: KKKKK (K) {20} KKKKK
 355: KKKKK (K) {20} KKKKK
 356: KKKKK (K) {20} KKKKK
 357: KKKKK (K) {20} KKKKK
 358: KKKKK (K) {20} KKKKK
 359: KKKKK (K) {20} KKKKK
 360: KKKKK (K) {20} KKKKK
 361: KKKKK (K) {20} KKKKK
 362: KKKKK (K) {20} KKKKK
 363: KKKKK (K) {20} KKKKK
 364: KKKKK (K) {20} KKKKK
 365: KKKKK (K) {20} KKKKK
 366: KKKKK (K) {20} KKKKK
 367: KKKKK (K) {20} KKKKK
 368: KKKKK (K) {20} KKKKK
 369: KKKKK (K) {20} KKKKK
 370: KKKKK (K) {20} KKKKK

81: KKKK (K){20} KKKKKKKKKKKKKKKKKKK

ABR01713 ck: 2112 len: 219 | ABr01713 Human breast specific polypeptide

(R,K){20,20}
(K){20}

190: PPPPP KKKKKKKKKKKKKKKKKKKKK KGGGG

191: PPPPP KKKKKKKKKKKKKKKKKKKKK KGGGT

192: PPPPK KKKKKKKKKKKKKKKKKKKKK GGGTS

AAE28374 ck: 5750 len: 20 | Aae28374 Peptide #1 used for transfection

(R,K){20,20}
(K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

AAE28375 ck: 7220 len: 20 | Aae28375 Peptide #1 used in the invention.

(R,K){20,20}
(R){20}

1: RRRRRRRRRRRRRRRRRRR

AAE28376 ck: 4440 len: 40 | Aae28376 Peptide #2 used in the invention.

(R,K){20,20}
(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKK KKKKK

2: K RKKKKKKKKKKKKKKKKKKKK KKKKK

3: KR KKKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK RKKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKR KKKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK RKKKKKKKKKKKKKKKKKKKK KKKKK

7: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKK RKKKKKKKKKKKKKKKKKKKK KKKKK

9: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKK RKKKKKKKKKKKKKKKKKKKK KKKKK

11: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKK RKKKKKKKKKKKKKKKKKKKK KKKKK

13: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}
14: KKKKK RKKKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}
15: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}
16: KKKKK RKKKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}
17: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}
18: KKKKK RKKKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}
19: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}
20: KKKKK RKKKKKKKKKKKKKKKKKKKK R

(R,K){20}
21: RKKKK KKKKKKKKKKKKKKKKKKKKK

AAE28379 ck: 598 len: 23 | Aae28379 RGD peptide #1 used in the invention

(R,K){20,20}
(K){20}

1: KKKKKKKKKKKKKKKKKKKKK RGD

(R,K){20}
2: K KKKKKKKKKKKKKKKKKKKKK GD

ABP66694 ck: 1944 len: 122 | ABP66694 Human breast specific protein SEQ

(R,K){20,20}
(K){20}

8: AARAG KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
9: ARAGK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
10: RAGKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
11: AGKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
12: GKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

30: IICLL (K){20} K
31: IICLL (K){20} K

ABG92605 ck: 5691 len: 108 ! Abg92605 Human DNA-binding protein #31. 11/

78: VRPCL (R,K){20,20} K
79: RPCLK (K){20} K
80: PCLKK (K){20} K
81: CLKKK (K){20} K
82: LKKKK (K){20} K
83: KKKKK (K){20} K
84: KKKKK (K){20} K
85: KKKKK (K){20} K
86: KKKKK (K){20} K
87: KKKKK (K){20} K
88: KKKKK (K){20} K
89: KKKKK (K){20} K

ABG92613 ck: 6029 len: 63 ! Abg92613 Human DNA-binding protein #39. 11/

40: KLTLK (R,K){20,20} ISWG

ABG92621 ck: 7170 len: 63 ! Abg92621 Human DNA-binding protein #47. 11/

37: TPSRA (R,K){20,20} K
38: PSRAK (K){20} K
39: SRAKK (K){20} K
40: RAKKK (K){20} K
41: AKKKK (K){20} K

42: KKKKK (K){20} K
43: KKKKK (K){20} K
44: KKKKK (K){20} K

ABG92625 ck: 6110 len: 61 ! Abg92625 Human DNA-binding protein #51. 11/

28: RPTRP (R,K){20,20} K
29: PTRPK (K){20} K
30: TRPKK (K){20} K
31: RPKKK (K){20} K
32: PKKKK (K){20} K
33: KKKKK (K){20} K
34: KKKKK (K){20} K
35: KKKKK (K){20} K
36: KKKKK (K){20} K
37: KKKKK (K){20} K

ABG92626 ck: 5764 len: 74 ! Abg92626 Human DNA-binding protein #52. 11/

40: EFLSA (R,K){20,20} K
41: FLSAK (K){20} K
42: LSACK (K){20} K
43: SAKKK (K){20} K
44: AKKKK (K){20} K
45: KKKKK (K){20} K
46: KKKKK (K){20} K
47: KKKKK (K){20} K
48: KKKKK (K){20} K

49: KKKKK (K){20} KKKKK
 50: KKKKK (K){20} KKKKK
 51: KKKKK (K){20} KKKKK
 52: KKKKK (K){20} KKKKK
 53: KKKKK (K){20} KKKKK
 54: KKKKK (K){20} KKKKK X

ABG92627 ck: 9217 len: 68 ! Abg92627 Human DNA-binding protein #53. 11/

38: FLFPE (R,K){20,20} KKKKK
 39: LFPEK (K){20} KKKKK
 40: FPEKK (K){20} KKKKK
 41: PEKKK (K){20} KKKKK
 42: EKKKK (K){20} KKKKK
 43: KKKKK (K){20} KKKKK
 44: KKKKK (K){20} KKKKK
 45: KKKKK (K){20} KKKKK

ABG92629 ck: 8162 len: 79 ! Abg92629 Human DNA-binding protein #55. 11/

41: VRPRV (R,K){20,20} KKKKK
 42: RPRVR (K){20} KKKKK
 43: PRVRK (K){20} KKKKK
 44: RVRKK (K){20} KKKKK
 45: VRKKK (K){20} KKKKK

ABG92658 ck: 285 len: 118 ! Abg92658 Human DNA-binding protein #84. 11/

98: EKHQK (R,K){20,20} KKKKK G

ABG92659 ck: 5509 len: 58 ! Abg92659 Human DNA-binding protein #85. 11/

36: FYFVC (R,K){20,20} KKKKK
 37: YFVCK (K){20} KKKKK
 38: FVCKK (K){20} KKKKK
 39: VCKKK (K){20} KKKKK

ABG92660 ck: 9074 len: 66 ! Abg92660 Human DNA-binding protein #86. 11/

40: LVQCE (R,K){20,20} KKKKK
 41: VQCEK (K){20} KKKKK
 42: QCEKK (K){20} KKKKK
 43: CEKKK (K){20} KKKKK
 44: EKKKK (K){20} KKKKK
 45: KKKKK (K){20} KKKKK
 46: KKKKK (K){20} KKKKK
 47: KKKKK (K){20} KKKKK

ABG92661 ck: 8528 len: 150 ! Abg92661 Human DNA-binding protein #87. 11/

113: SRNTV (R,K){20,20} KKKKK
 114: RNTVK (K){20} KKKKK
 115: NTVKK (K){20} KKKKK
 116: TVKKK (K){20} KKKKK
 117: VKKKK (K){20} KKKKK
 118: KKKKK (K){20} KKKKK
 119: KKKKK (K){20} KKKKK
 120: KKKKK (K){20} KKKKK
 121: KKKKK (K){20} KKKKK

91: KKKKK (K){20} K
92: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK
ABG92668 ck: 8102 len: 111 ! Abg92668 Human DNA-binding protein #94. 11/
(R,K){20,20}
78: EPHIL KKKKKKKKKKKKKKKKKKKKKKKKKKK
79: PHILK KKKKKKKKKKKKKKKKKKKKKKKKKKK
80: HILKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
81: ILKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
82: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
83: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
84: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
85: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
86: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
87: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
88: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
89: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
90: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
91: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
92: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

ABG92669 ck: 8319 len: 53 ! Abg92669 Human DNA-binding protein #95. 11/
(R,K){20,20}
13: RYFKP KKKKKKKKKKKKKKKKKKKKKKKKKKK
14: YPKPK KKKKKKKKKKKKKKKKKKKKKKKKKKK
15: FKPKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
16: KPKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
17: PKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK

1

18: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
19: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
20: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
25: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
26: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
27: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
28: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
30: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
ABG92671 ck: 7918 len: 80 ! Abg92671 Human DNA-binding protein #97. 11/
(R,K){20,20}
50: NVLTV KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
51: VLTVR KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
52: LTVKR KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
53: TVKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
54: VKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
55: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
56: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK

1

ABG92673 ck: 4882 len: 41 ! Abg92673 Human DNA-binding protein #99. 11/
(R,K){20,20}
8: FYCFF KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

1

9: YCFEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
10: CFFKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
11: PFKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
12: FKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

ABG92674 ck: 5469 len: 63 | Abg92674 Human DNA-binding protein #100. 11

30: IICLL KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20,20}
(K) {20}
31: ICLLK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
32: CLLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
33: LLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
34: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
35: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
ABG92675 ck: 5075 len: 52 | Abg92675 Human DNA-binding protein #101. 11
(R,K) {20,20}
(K) {20}
30: FIVVK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
31: IVVVK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
32: VVVKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
33: VVKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

ABG92676 ck: 5741 len: 47 | Abg92676 Human DNA-binding protein #102. 11

20: ILTTF KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20,20}
(K) {20}
21: LTTFK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
22: TTFKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
23: TFKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
24: FKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
26: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

ABG92677 ck: 2869 len: 84 | Abg92677 Human DNA-binding protein #103. 11

53: KCTVE KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20,20}
(K) {20}
54: CTVEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
55: TYEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
56: YEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
57: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

[illegible]

ABG92678 ck: 4686 len: 73 ! Abg92678 Human DNA-binding protein #104. 11

[illegible]

ABG92679 ck: 6676 len: 74 ! Abg92679 Human DNA-binding protein #105. 11

```

47: LRTEQ KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (R,K){20,20}
      (K){20}

48: RTFOK KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}

49: TFOKK KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}

50: FOKKK KKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
```

[illegible]

ABG92680 ck: 2283 len: 54 ! Abg92680 Human DNA-binding protein #106. 1:

```

32: IVFCF KKKKKKKKKKKKKKKKKKKKK KKK
      (R, K) {20, 20}
      (K) {20}

33: VFCPK KKKKKKKKKKKKKKKKKKKKK KX
      (K) {20}

34: FCFKK KKKKKKKKKKKKKKKKKKKKK X
      (K) {20}

```

ABG92681 ck: 7503 len: 74 ! Abg92681 Human DNA-binding protein #107. 1

```

45: SHLTD (R, K) { 20, 20 }
      (K) { 20 }

46: HLTDK (K) { 20 }

47: LTDKK (K) { 20 }

48: TDKKK (K) { 20 }

49: DKKKK (K) { 20 }

50: KKKKK (K) { 20 }

51: KKKKK (K) { 20 }

52: KKKKK (K) { 20 }

53: KKKKK (K) { 20 }

54: KKKKK (K) { 20 }

55: KKKKK (K) { 20 }

```

ABG92683 ck: 5199 len: 84 ! Abg92683 Human DNA-binding protein #109. 1:

(R,K) { 20, 20 }
 (K) { 20 }

63: AYNAS KKKKKKKKKKKKKKKKKKKKK XG

ABG92684 ck: 7578 len: 31 | Abg92684 Human DNA-binding protein #110. 1:

[illegible]

8: ELEKK KKKKKKKKKKKKKKKKKKKKK KKKX
(K) {20}
9: LEKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K) {20}
10: EKKKK KKKKKKKKKKKKKKKKKKKKK KX
(K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK X

ABG92685 ck: 3915 len: 57 ! Abg92685 Human DNA-binding protein #111. 11

(R,K) {20,20}
31: KQLLL KKKKKKKKKKKKKKKKKKKKK KKKXG
(K) {20}
32: QLLLK KKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
33: LLLKK KKKKKKKKKKKKKKKKKKKKK KXGGF
(K) {20}
34: LLKKK KKKKKKKKKKKKKKKKKKKKK XGGF

ABG92686 ck: 3679 len: 37 ! Abg92686 Human DNA-binding protein #112. 11

(R,K) {20,20}
15: ISPLT KKKKKKKKKKKKKKKKKKKKK KKK
(K) {20}
16: SPLTK KKKKKKKKKKKKKKKKKKKKK KX
(K) {20}
17: PLTKK KKKKKKKKKKKKKKKKKKKKK X

ABG92687 ck: 657 len: 196 ! Abg92687 Human DNA-binding protein #113. 11

(R,K) {20,20}
169: FVXFE KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
170: VXFEX KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
171: XFEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
172: FEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
173: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKX
(K) {20}
174: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K) {20}
175: KKKKK KKKKKKKKKKKKKKKKKKKKK KX
(K) {20}
176: KKKKK KKKKKKKKKKKKKKKKKKKKK X

ABG92688 ck: 4672 len: 57 ! Abg92688 Human DNA-binding protein #114. 11

(R,K) {20,20}
(K) {20}

28: DKTFH KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
29: KTFHK KKKKKKKKKKKKKKKKKKKKK KKKXP
(K) {20}
30: TFHKK KKKKKKKKKKKKKKKKKKKKK KXPPG
(K) {20}
31: FHKKK KKKKKKKKKKKKKKKKKKKKK KXPGG
(K) {20}
32: HKKKK KKKKKKKKKKKKKKKKKKKKK XPGGG

ABG92689 ck: 9656 len: 66 ! Abg92689 Human DNA-binding protein #115. 11

(R,K) {20,20}
38: MVISV KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
39: VISVK KKKKKKKKKKKKKKKKKKKKK KKKKE
(K) {20}
40: ISVKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
41: SVKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
42: VKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K) {20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK

ABG92691 ck: 4665 len: 34 ! Abg92691 Human DNA-binding protein #117. 11

(R,K) {20,20}
10: PELLK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
11: ELLKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
12: LLLKK KKKKKKKKKKKKKKKKKKKKK KKK
(K) {20}
13: LLKKK KKKKKKKKKKKKKKKKKKKKK KK
(K) {20}
14: LKKKK KKKKKKKKKKKKKKKKKKKKK K
(K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK

ABG92692 ck: 7810 len: 64 ! Abg92692 Human DNA-binding protein #118. 11

(R,K) {20,20}
37: LKYPW KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
38: KYFWK KKKKKKKKKKKKKKKKKKKKK KKKXG
(K) {20}
39: YFWKK KKKKKKKKKKKKKKKKKKKKK KKKXG
(K) {20}
40: FWKKK KKKKKKKKKKKKKKKKKKKKK KXGXP

41: WKXXX (K){20}
ABG92693 ck: 269 len: 76 ! Abg92693 Human DNA-binding protein #119. 11
37: TPSRA (R,K){20,20}
(K){20}
38: PSRAK (K){20}
39: SRAKK (K){20}
40: RAKKK (K){20}
41: AKKKK (K){20}
42: KKKKK (K){20}
43: KKKKK (K){20}
44: KKKKK (K){20}

ABG92694 ck: 8370 len: 45 ! Abg92694 Human DNA-binding protein #120. 11

17: APKTQ (R,K){20,20}
(K){20}
18: PKTQK (K){20}
19: KTQKK (K){20}
20: TQKKK (K){20}
21: QKXXX (K){20}
22: KKKKK (K){20}
23: KKKKK (K){20}
24: KKKKK (K){20}
25: KKKKK (K){20}

ABG92695 ck: 1663 len: 87 ! Abg92695 Human DNA-binding protein #121. 11

36: KWSSX (R,K){20,20}
(K){20}
37: WSSXX (K){20}

38: SSXXX (K){20}
39: SXXXX (K){20}
40: XXXXX (K){20}
41: XXXXX (K){20}
42: XXXXX (K){20}
43: XXXXX (K){20}
44: XXXXX (K){20}
45: XXXXX (K){20}
46: XXXXX (K){20}
47: XXXXX (K){20}
48: XXXXX (K){20}
49: XXXXX (K){20}
50: XXXXX (K){20}
51: XXXXX (K){20}
52: XXXXX (K){20}
53: XXXXX (K){20}
54: XXXXX (K){20}
55: XXXXX (K){20}
56: XXXXX (K){20}
57: XXXXX (K){20}
58: XXXXX (K){20}
59: XXXXX (K){20}
60: XXXXX (K){20}
61: XXXXX (K){20}

127: **XXXXXXXXXXXXXXXXXXXXXXXXXXXX** **(K) {20}**

81: VLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKK KKK
(K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKK K
(K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKK

ABG65251 ck: 8971 len: 108 | Abg65251 Human albumin fusion protein #1926

1

(R,K) {20,20}
(K) {20}
78: TLXLK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
79: LLXLK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
80: LXLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
81: XLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

ABG65440 ck: 1431 len: 530 | Abg65440 Human albumin fusion protein #2115

1

(R,K) {20,20}
(K) {20}
511: LHAPP KKKKKKKKKKKKKKKKKKK

ABG35896 ck: 5383 len: 86 | Abg35896 Human peptide encoded by genome-de

1

(R,K) {20,20}
(R,K) {20}
15: RRRRG RRRKKKKKKKKKKKKKKKK KKKRR
(R,K) {20}
16: RRRGR RRRKKKKKKKKKKKKKKKK KRRRR
(R,K) {20}
17: RRGRR RRRKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
18: RGRRR RRRKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}

19: GRRRR RKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
20: RRRRR KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
21: RRRRK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
22: RRRKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
23: RRJJK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
24: RKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
25: KKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
26: KKKKK RKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
27: KKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
28: KKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
29: KKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
30: KKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
31: RKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
32: KKKKK RKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
33: KKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
34: KKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
35: KKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R,K) {20}
36: KKKKK KKKKKKKKKKKKKKKKKKK RRRRR
(R) {20}
37: RKKKK RRRRRKKKKKKKKKKKKKK RRRRR
(R) {20}
38: KKKKK RRRRRKKKKKKKKKKKKKK RRRRR
(R) {20}
39: KKKRR RRRRRKKKKKKKKKKKKKK RRRRR
(R) {20}
40: KKKRR RRRRRKKKKKKKKKKKKKK RRRRR
(R) {20}
41: KRRRR RRRRRKKKKKKKKKKKKKK RRRRR
(R) {20}
42: RRRRR RRRRRKKKKKKKKKKKKKK RRRRR
(R) {20}
43: RRRRR RRRRRKKKKKKKKKKKKKK RRRRR
(R) {20}

44: RRRR RRRRRRRRRRRRRRRR RNKQT
(R){20}
45: RRRR RRRRRRRRRRRRRRRR NKQTK
(R){20}
ABG36760 ck: 1334 len: 86 ! Abg36760 Human peptide encoded by genome-de
(R,K){20,20}
57: EEEG RRRKKRRKKKKKKKKKK KKKK
(R,K){20}
58: EEEG RRRKKRRKKKKKKKKKK KKKK
(R,K){20}
59: EGGRR RKKKKRRKKKKKKKKKK KKKK
(R,K){20}
60: EGGRR RKKKKRRKKKKKKKKKK KKKK
(R,K){20}
61: GRRR KKKRRKKKKKKKKKKKK KKKK
(R,K){20}
62: RRRK KRRKKKKKKKKKKKKKK KKKK
(R,K){20}
63: RRRK KRRKKKKKKKKKKKKKK KKKK
(R,K){20}
64: RKKK RRRKKKKKKKKKKKKKK KKKK
(R,K){20}
65: KKKK KRRKKKKKKKKKKKKKK KKKK
(R,K){20}
66: KKKK RRRKKKKKKKKKKKKKK KKKK
(R,K){20}
67: KRRK KRRKKKKKKKKKKKKKK KKKK
(K){20}

ABG36843 ck: 9082 len: 167 ! Abg36843 Human peptide encoded by genome-de
(R,K){20,20}
(R,K){20}
33: EEGG RRRRRRRRRRRRRRRR RGGG
(R,K){20}
34: EGGRR RRRRRRRRRRRRRRRR RGGG
(R,K){20}
35: GGGRR RRRRRRRRRRRRRRRR GGGRR
(R,K){20}

ABG37280 ck: 2276 len: 89 ! Abg37280 Human peptide encoded by genome-de
(R,K){20,20}
(R,K){20}
23: EEEE KRRKKRRKKKKKKKKKK EEEK
(K){20}
46: KKEE KRRKKRRKKKKKKKKKK KKEE
(K){20}
47: KEEK KRRKKRRKKKKKKKKKK KKEE
(R,K){20}
48: EEEK KRRKKRRKKKKKKKKKK KKEE

49: EEEK (R,K){20}
ABG37848 ck: 3607 len: 88 ! Abg37848 Human peptide encoded by genome-de
(R,K){20,20}
(K){20}
39: ERRE KRRKKRRKKKKKKKKKK KKKK
(K){20}
40: RREK KRRKKRRKKKKKKKKKK KKKK
(K){20}
41: RREK KRRKKRRKKKKKKKKKK KKKK
(K){20}
42: RREK KRRKKRRKKKKKKKKKK KKKK
(K){20}
43: EKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
44: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
45: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
46: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
47: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
48: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
49: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
50: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
51: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
52: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
53: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
54: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
55: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
56: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
57: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
58: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
59: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}
60: KKKK KRRKKRRKKKKKKKKKK KKKK
(K){20}

[illegible]

1

46: KKKK (R,K){20} KKKKKKKKKKKKKKK KKS
47: KKKR (K){20} KKKKKKKKKKKKKKKKK KKS
48: KKKK (K){20} KKKKKKKKKKKKKKKKK KSAH
49: KKKK (K){20} KKKKKKKKKKKKKKKKK SAH

ABG40383 ck: 1560 len: 88 ! Abg40383 Human peptide encoded by genome-de

43: RRRG (R,K){20,20} (R){20} RRRRRRRRRRRRRRRRR RRRR
44: RRRG (R){20} RRRRRRRRRRRRRRRRR RRRR
45: RRRG (R){20} RRRRRRRRRRRRRRRRR RRRR
46: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
47: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
48: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
49: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
50: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
51: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
52: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
53: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
54: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
55: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
56: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
57: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
58: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
59: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
60: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
61: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR

62: RRRR (R){20} RRRRRRRRRRRRRRRRR NTNN
ABG40676 ck: 2324 len: 36 ! Abg40676 Human peptide encoded by genome-de
16: KRRK (R,K){20,20} (R,K){20} KRRRRRRRRRRRRRRRR R
17: ERRT (R,K){20} KRRRRRRRRRRRRRRRR

ABG43651 ck: 8343 len: 66 ! Abg43651 Human peptide encoded by genome-de

6: ETERE (R,K){20,20} (R,K){20} KKKKKKKKKKKKKKK KKKK
7: TEREK KKKKKKKKKKKKKKKKKKKKK KKKK
8: EREKK KKKKKKKKKKKKKKKKKKKKK KKKK
9: REKKK KKKKKKKKKKKKKKKKKKKKK KKKK

ABG45640 ck: 2394 len: 57 ! Abg45640 Human peptide encoded by genome-de

20: EEEG (R,K){20,20} (R){20} RRRRRRRRRRRRRRRRR RRRR
21: EEEG (R){20} RRRRRRRRRRRRRRRRR RRRR
22: EEEG (R){20} RRRRRRRRRRRRRRRRR RRRR
23: EEEG (R){20} RRRRRRRRRRRRRRRRR RRRR
24: GRRR (R){20} RRRRRRRRRRRRRRRRR RRRR
25: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRR

ABG46635 ck: 3301 len: 52 ! Abg46635 Human peptide encoded by genome-de

12: KKKK (R,K){20,20} (R,K){20} KKKKKKKKKKKKKKK KKKK
13: KKKK (R,K){20} KKKKKKKKKKKKKKKKK KKKK
14: KKKK (R,K){20} KKKKKKKKKKKKKKKKK KKKK
15: KKKK (R,K){20} KKKKKKKKKKKKKKKKK KKKK
16: KKKK (R,K){20} KKKKKKKKKKKKKKKKK KKKK
17: KKKK (R,K){20} KKKKKKKKKKKKKKKKK KKKK

18: KKKKK (R,K){20} KKKKK
19: KKKKK (R,K){20} KKKKK
20: KKKKK (R,K){20} KKKKK
21: KKKKK (R,K){20} KKKKK
22: KKKKK (R,K){20} KKKKK
23: KKKKK (R,K){20} KKKKK
24: KKKKK (R,K){20} KKKKK
25: KKKKK (R,K){20} KKKKK
26: KKKKK (R,K){20} KKKKK
27: KKKKK (R,K){20} KKKKK
28: KKKKK (R,K){20} KKKKK
29: KKKKK (R,K){20} KKKKK
30: KKKKK (R,K){20} KKKKK
31: KKKKK (R,K){20} KKKKK

ABG47062 ck: 4895 len: 51 ! Abg47062 Human peptide encoded by genome-de
18: LFKPM KKKKKKKKKKKKKKKKKKK KCLTT
19: FKPMR KKKKKKKKKKKKKKKKKKK KLTFT
20: KPMRK KKKKKKKKKKKKKKKKKKK LTTTT

ABB77424 ck: 9633 len: 637 ! Abb77424 Human tumour marker protein seq2-5.
618: GDKTD KKKKKKKKKKKKKKKKKKK

AAE20631 ck: 8971 len: 108 ! Aae20631 Human gene 4 encoded secreted prob
78: TLLXL KKKKKKKKKKKKKKKKKKK KKKKK
79: LLLK KKKKKKKKKKKKKKKKKKK KKKKK

80: LXLKK KKKKKKKKKKKKKKKKKKK KKKKK
81: XLKKK (K){20} KKKKK
82: LKKKK (K){20} KKKKK
83: KKKKK (K){20} KKKKK
84: KKKKK (K){20} KKKKK
85: KKKKK (K){20} KKKKK
86: KKKKK (K){20} KKKKK

AAE20642 ck: 7046 len: 105 ! Aae20642 Human gene 4 encoded secreted pro

78: TLLVL (R,K){20,20} KKKKK
79: LLVLK (K){20} KKKKK
80: LVLKK (K){20} KKKKK
81: VLKKK (K){20} KKKKK
82: LKKKK (K){20} KKKKK
83: KKKKK (K){20} KKKKK
84: KKKKK (K){20} KKKKK
85: KKKKK (K){20} KKKKK
86: KKKKK (K){20} KKKKK

ABB89690 ck: 1158 len: 226 ! Abb89690 Human polypeptide SEQ ID NO 2066.
198: GAESL (R,K){20,20} KKKKK
199: AESLK (K){20} KKKKK
200: ESLKK (K){20} KKKKK

AAE14544 ck: 7250 len: 35 ! Aae14544 Peptide p65 used in assay for det
1: (R,K){20,20} KKKKK
2: (K){20} KKKKK

3: KK (K){20} KKKK
4: KK (K){20} KKKK
5: KK (K){20} KKKK
6: KK (K){20} KKKK
7: KK (K){20} KKKK
8: KK (K){20} KKKK
9: KK (K){20} KKKK
10: KK (K){20} KKKK
11: KK (K){20} KKKK
12: KK (K){20} KKKK
13: KK (K){20} KKKK
14: KK (K){20} KKKK
15: KK (K){20} K
16: KK (K){20} KKKK
AAU75162 ck: 5536 len: 84 ! Aau75162 Single-chain antigen-binding polyh
(R,K){20,20}
1: KKKK
2: K KKKK
3: KK (K){20} KKKK
4: KK (K){20} KKKK
5: KK (K){20} KKKK
6: KK (K){20} KKKK
7: KK (K){20} KKKK
8: KK (K){20} KKKK
9: KK (K){20} KKKK

10: KK (K){20} KKKK
11: KK (K){20} KKKK
12: KK (K){20} KKKK
13: KK (K){20} KKKK
14: KK (K){20} KKKK
15: KK (K){20} KKKK
16: KK (K){20} KKKK
17: KK (K){20} KKKK
18: KK (K){20} KKKK
19: KK (K){20} KKKK
20: KK (K){20} KKKK
21: KK (K){20} KKKK
22: KK (K){20} KKKK
23: KK (K){20} KKKK
24: KK (K){20} KKKK
25: KK (K){20} KKKK
26: KK (K){20} KKKK
27: KK (K){20} KKKK
28: KK (K){20} KKKK
29: KK (K){20} KKKK
30: KK (K){20} KKKK
31: KK (K){20} KKKK
32: KK (K){20} KKKK
33: KK (K){20} KKKK

34: KKKKK KKKKKKKKKKKKKKKKKKK KKKXX
(K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKXX
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKXX
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKXX
(K){20}

AAU75163 ck: 8024 len: 84 ! Aau75163 Single-chain antigen-binding poly

1

(R,K){20,20}
(R){20}
1: RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
2: R RRRRRRRRRRRRRRRRR RRRRR
(R){20}
3: RR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
4: RRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
5: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
6: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
7: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
8: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
9: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
10: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
11: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
12: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
13: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
14: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
15: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
16: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
17: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
18: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
19: RRRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}

(R){20}
20: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
21: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
22: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
23: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
24: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
25: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
26: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
27: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
28: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
29: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
30: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
31: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
32: RRRR RRRRRRRRRRRRRRRRR RRRRR
(R){20}
33: RRRR RRRRRRRRRRRRRRRRR RRRRX
(R){20}
34: RRRR RRRRRRRRRRRRRRRRR RRRXX
(R){20}
35: RRRR RRRRRRRRRRRRRRRRR RRRXX
(R){20}
36: RRRR RRRRRRRRRRRRRRRRR RRRXX
(R){20}
37: RRRR RRRRRRRRRRRRRRRRR RRRXX
(R){20}

AAU75164 ck: 9643 len: 83 ! Aau75164 Single-chain antigen-binding poly

1

(R,K){20,20}
(R,K){20}
1: RRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
2: R RRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
3: RR RRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
4: RRR RRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
5: RRR RRRRRRRRRRRRRRRRR RRRRR
(R,K){20}

16: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 17: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 18: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 19: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 20: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 21: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 22: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 23: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 24: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 25: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 26: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 27: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 28: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 29: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 30: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 31: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 32: RRRR RRRRRRRRRRRRRRRRRRRR RRRR
 (R) {20}
 33: RRRR RRRRRRRRRRRRRRRRRRRR RRRX
 (R) {20}
 34: RRRR RRRRRRRRRRRRRRRRRRRR RRRX
 (R) {20}
 35: RRRR RRRRRRRRRRRRRRRRRRRR RRRX
 (R) {20}
 36: RRRR RRRRRRRRRRRRRRRRRRRR RRRX
 (R) {20}
 37: RRRR RRRRRRRRRRRRRRRRRRRR RRRX
 (R) {20}

AAE16634 ck: 1504 len: 32 ! Aae16634 Nuclear localisation signal peptid

1:
 (R,K) {20,20}
 (R,K) {20}
 KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}

2: K KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}
 3: KR KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}
 4: KKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}
 5: KKKR KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}
 6: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}
 7: KKKR KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}
 8: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}
 9: KKKR KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}
 10: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}
 11: KKKR KKKKKKKKKKKKKKKKKKKKK KKKK
 (R,K) {20}
 12: KKKK KKKKKKKKKKKKKKKKKKKKK R
 (R,K) {20}
 13: KKKR KKKKKKKKKKKKKKKKKKKKK

AA448798 ck: 6221 len: 26 ! Aam48798 Tumour-targetting peptide vector

1
 (R,K) {20,20}
 (K) {20}
 7: RGDGF KKKKKKKKKKKKKKKKKKKKK

ABB44830 ck: 5275 len: 38 ! Abb44830 Human protective sequence CNI-007

1
 (R,K) {20,20}
 (K) {20}
 4: MAQ KKKKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 5: MAQK KKKKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 6: MAQKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 7: AQKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 8: QKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}
 12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (K) {20}

13: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKG
 14: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKGG
 15: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKGG
 16: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KGG
 17: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK GG

AAU69690 ck: 875 len: 46 ! Aau69690 Cell death protective sequence CNP
 (R,K){20,20}
 (K){20}

15: REKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 16: EKSKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 17: KSKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 18: KSKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 19: SKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 20: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 21: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 22: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 23: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 24: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KGG
 25: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK GG

AAU69736 ck: 5862 len: 50 ! Aau69736 Cell death protective sequence CNP
 (R,K){20,20}
 (R,K){20}

11: ESALG KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 12: SALGR KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 13: ALGRK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 14: LGRKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 15: GRKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 16: RKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 18: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 19: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 20: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 25: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 26: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 27: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
 28: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KGR
 29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKKKK GR

ABG47914 ck: 5383 len: 86 ! Abg47914 Human liver peptide, SEQ ID No 265
 (R,K){20,20}
 (R,K){20}

15: RRRRG RRRRRKKKKKKKKKKKKKKKKKKKKKK KKKRR
 16: RRRGR RRRKKKKKKKKKKKKKKKKKKKKKK KRRRR
 17: RRRGR RRRKKKKKKKKKKKKKKKKKKKKKK RRRRR
 18: RRRRR RRRKKKKKKKKKKKKKKKKKKKKKK RRRRR
 19: GRRRR RRRKKKKKKKKKKKKKKKKKKKKKK RRRRR
 20: RRRRR RRRKKKKKKKKKKKKKKKKKKKKKK RRRRR
 21: RRRRK RRRKKKKKKKKKKKKKKKKKKKKKK RRRRR
 22: RRRKK RRRKKKKKKKKKKKKKKKKKKKKKK RRRRR
 23: RRRKK RRRKKKKKKKKKKKKKKKKKKKKKK RRRRR
 24: RRRKK RRRKKKKKKKKKKKKKKKKKKKKKK RRRRR

45:	KERR	RRRRRRRRRRRRRRRRRRR	RRRKX (R,K){20}
46:	ERRR	RRRRRRRRRRRRRRRRR	RRKKK (R,K){20}
47:	RRRR	RRRRRRRRRRRRRRRRR	RKKKK (R,K){20}
48:	RRRR	RRRRRRRRRRRRRRRRR	KKKKK (R,K){20}
49:	RRRR	RRRRRRRRRRRRRRRRR	KKKKK (R,K){20}
50:	RRRR	RRRRRRRRRRRRRRRRR	KKKKK (R,K){20}
51:	RRRR	RRRRRRRRRRRRRRRRR	KKKKE (R,K){20}
52:	RRRR	RRRRRRRRRRRRRRRRR	KKEEE (R,K){20}
53:	RRRR	RRRRRRRRRRRRRRRRR	KEEEE (R,K){20}
54:	RRRR	RRRRRRRRRRRRRRRRR	EEEEE (R,K){20}

ABG49962 ck: 3607 len: 88 ! Abg49962 Human liver peptide, SEQ ID No 286

[illegible]

```

52: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
53: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
54: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
55: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
56: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
57: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
58: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
59: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
60: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
61: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
62: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
63: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
64: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
65: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
66: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
67: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
68: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
69: KKKKK (K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK

```

ABG50530 ck: 3937 len: 85 ! Abg50530 Human liver peptide, SEQ ID No 291

1 :	$(R, K) \{20, 20\}$	$\{\begin{smallmatrix} R \\ K \end{smallmatrix}\} \{20\}$
2 :	$K \{20\}$	$\{K\} \{20\}$
3 :	$K \{20\}$	$\{K\} \{20\}$
4 :	$K \{20\}$	$\{K\} \{20\}$
5 :	$K \{20\}$	$\{K\} \{20\}$

6: KKKK KKKKKKKKKKKKKKKKKKK KKKK
7: KKKK KKKKKKKKKKKKKKKKKKK KKKK
8: KKKK KKKKKKKKKKKKKKKKKKK KKKK
9: KKKK KKKKKKKKKKKKKKKKKKK KKKK
10: KKKK KKKKKKKKKKKKKKKKKKK KKKK
11: KKKK KKKKKKKKKKKKKKKKKKK KKKK
ABG52382 ck: 1560 len: 88 ! Abg52382 Human liver peptide, SEQ ID No 310
43: RRRG RRRRRRRRRRRRRRRRRR RRRR
44: RRRG RRRRRRRRRRRRRRRRRR RRRR
45: RRRG RRRRRRRRRRRRRRRRRR RRRR
46: RRRG RRRRRRRRRRRRRRRRRR RRRR
47: RRRR RRRRRRRRRRRRRRRRRR RRRR
48: RRRR RRRRRRRRRRRRRRRRRR RRRR
49: RRRR RRRRRRRRRRRRRRRRRR RRRR
50: RRRR RRRRRRRRRRRRRRRRRR RRRR
51: RRRR RRRRRRRRRRRRRRRRRR RRRR
52: RRRR RRRRRRRRRRRRRRRRRR RRRR
53: RRRR RRRRRRRRRRRRRRRRRR RRRR
54: RRRR RRRRRRRRRRRRRRRRRR RRRR
55: RRRR RRRRRRRRRRRRRRRRRR RRRR
56: RRRR RRRRRRRRRRRRRRRRRR RRRR
57: RRRR RRRRRRRRRRRRRRRRRR RRRR
58: RRRR RRRRRRRRRRRRRRRRRR RRRR
59: RRRR RRRRRRRRRRRRRRRRRR RRRR

60: RRRR RRRRRRRRRRRRRRRRRR RRRR
61: RRRR RRRRRRRRRRRRRRRRRR RRRR
62: RRRR RRRRRRRRRRRRRRRRRR RRRR
ABG52597 ck: 2324 len: 36 ! Abg52597 Human liver peptide, SEQ ID No 31:
16: KERK KKKKKKKKKKKKKKKKKKK R
17: ERKTK KKKKKKKKKKKKKKKKKKK
ABG55512 ck: 8343 len: 66 ! Abg55512 Human liver peptide, SEQ ID No 34:
6: ETERE KKKKKKKKKKKKKKKKKKK KKKK
7: TEREK KKKKKKKKKKKKKKKKKKK KKKK
8: EREK KKKKKKKKKKKKKKKKKKK KKKK
9: REKK KKKKKKKKKKKKKKKKKKK KKKK
ABG58055 ck: 2394 len: 57 ! Abg58055 Human liver peptide, SEQ ID No 36:
20: EEEG RRRRRRRRRRRRRRRRRR RRRR
21: EEEG RRRRRRRRRRRRRRRRRR RRRR
22: EEEG RRRRRRRRRRRRRRRRRR RRRR
23: EEEG RRRRRRRRRRRRRRRRRR RRRR
24: GRRR RRRRRRRRRRRRRRRRRR RRRR
25: RRRR RRRRRRRRRRRRRRRRRR RRRR
ABG58577 ck: 4228 len: 24 ! Abg58577 Human liver peptide, SEQ ID No 37:
1: RRRRRRRRRRRRRRRRRRR RRRR
2: RRRRRRRRRRRRRRRRRRR RRRR
3: RR RRRRRRRRRRRRRRRRR RRRR
4: RRR RRRRRRRRRRRRRRRRR RRRR

1

ABG59254 ck: 3301 len: 52 ! Abg59254 Human liver peptide, SEQ ID No 379

(R,K){20,20}

(R,K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

ABG59685 ck: 4895 len: 51 ! Abg59685 Human liver peptide, SEQ ID No 383

(R,K){20,20}

(R,K){20}

18: LFKPM KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

19: FKPMR KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

20: KPMRK RRRKKKKKKKKKKKKKKKK LTTTT

ABB96057 ck: 5626 len: 139 ! Abb96057 Human testicular antigen SEQ ID N

(R,K){20,20}

(K){20}

111: IHLNL KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

112: HLNLK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

113: LNLKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

114: NLKKK KKKKKKKKKKKKKKKKKKK KKKKK

ABB96575 ck: 4751 len: 80 ! Abb96575 Human testicular antigen SEQ ID N

(R,K){20,20}

(K){20}

61: KKKFD KKKKKKKKKKKKKKKKKKK KKKKK

AAU87124 ck: 5599 len: 281 ! Aau87124 Novel central nervous system prote

(R,K){20,20}

(K){20}

248: KKKKQ KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

249: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

250: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

251: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

252: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

253: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

254: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

255: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

256: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

257: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

258: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

259: KKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

260: KKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

261: KKKKK KKKKKKKKKKKKKKKKKKK K

262: KKKKK KKKKKKKKKKKKKKKKKKKKKK
 (κ) {20}

ABG00401 ck: 5116 len: 1,074 | Abg00401 Novel human diagnostic protein #39

```

(R, K) {20, 20}
(R, K) {20}
609: RGSSS KKKKKKKKKKKKKKKRR KKKKK

```

```
610: GSSK KKKRKKKKKKKKKKRKR NRKKK
      (R, K) {20}
```

ABG03974 ck: 5341 len: 99 ! Abg03974 Novel human diagnostic protein #399

$(R, K) \{20, 20\}$

[illegible]

4: MCK KKKKKKKKKKKKKKKKKKKKKKKK XNISO
 (K) {20}

5: MKKK KKKKKKKKKKKKKKKKKKKKK KNSOI
(K) {20}

```

6: MKKKK KKKKKKKKKKKKKKKKKKKKK NSQID
      (K) {20}

```

ABG04391 ck: 9047 len: 139 ! Abg04391 Novel human diagnostic protein #43

(R,K) {20,20}
(K) {20}

79: EEEEE KKKKKKKKKKKKKKKKKKK

80: EEEEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

81: EEEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

82: EEKKK KKKKKKKKKKKKKKKKKKKKKKKKK

[illegible][illegible]

85: $(K) \{20\}$

[illegible]

87: $\mathbb{K} \{ \mathbb{K} \}^{20}$

[illegible]

89: $(K) \{20\}$

[illegible]

(K) {20}

ABG05352 ck: 1276 len: 204 ! AbG05352 Novel human diagnostic protein #5.
 (R,K){20,20}
 (K){20}

106: EEEEE KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
107: EEEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
108: EEEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
109: EEEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
110: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
111: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
112: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
113: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
114: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
115: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
116: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
117: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
118: KKKKK KKKKKKKKKKKKKKKKKKK EEEEE
(R, K) {20}

ABG05367 ck: 6907 len: 808 | Abg05367 Novel human diagnostic protein #53

219: EEEEE RKKKKKKKKKKKKKKKKKKK RRRRR
(R, K) {20, 20}
(R, K) {20}
220: EEEER RKKKKKKKKKKKKKKKKKKK RRRRR
(R, K) {20}
221: EEEER KKKKKKKKKKKKKKKKKKK RRRKK
(R, K) {20}
222: EERRK KKKKKKKKKKKKKKKKKKK RRRKK
(R, K) {20}
223: ERRKK KKKKKKKKKKKKKKKKKKK RKKKK
(R, K) {20}
224: RKKKK KKKKKKKKKKKKKKKKKKK RKKKK
(R, K) {20}
225: RKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
226: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
227: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
228: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}

229: KKKKK RRRRRRRRRRRRRRRRRKK KKKKK
(R, K) {20}
230: KKKKK RRRRRRRRRRRRRRRRRKK KKKKK
(R, K) {20}
231: KKKKK RRRRRRRRRRRRRRRRRKK KKKKK
(R, K) {20}
232: KKKKK RRRRRRRRRRRRRRRRRKK KKKKK
(R, K) {20}
233: KRRRR RRRRRRRRRRRRRRRRRKK KKKKK
(R, K) {20}
234: RRRRR RRRRRRRRRRRRRRRRRKK KKKKK
(R, K) {20}
235: RRRRR RRRRRRRRRRRRRRRRRKK KKKKK
(R, K) {20}
236: RRRRR RRRRRRRRRRRRRRRRRKK KKKKK
(R, K) {20}
237: RRRRR RRRRRRRRRRRRRRRRRKK KKKKK
(K) {20}
258: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
259: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
260: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

ABG06375 ck: 7807 len: 2,570 | Abg06375 Novel human diagnostic protein #63

(R, K) {20, 20}
(R, K) {20}
205: KKKSS RRRRRRRRRRRRRRRRRRR RRRQQ
(R, K) {20}
206: KSSSR RRRRRRRRRRRRRRRRRRR RRRQQ
(R, K) {20}
207: KSSRR RRRRRRRRRRRRRRRRRRR RRRQQ
(R, K) {20}
208: SSRRR RRRRRRRRRRRRRRRRRRR RRRQQ
(R, K) {20}

ABG06513 ck: 2934 len: 154 | Abg06513 Novel human diagnostic protein #65

(R, K) {20, 20}
(R, K) {20}
49: KKKKT RKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
50: KKKTR KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
51: KKTAK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
52: KTRKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
53: TRKKR KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}
54: RKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R, K) {20}

1

1

1

55: KKKKK (R,K){20} KKKKK
56: KKKKK (R,K){20} KKKKK
57: KKKKK (R,K){20} KKKKK
58: KKKKK (R,K){20} KKKKK
59: KKKKK (R,K){20} KKKKK
60: KKKKK (K){20} KKKKK
61: KKKKK (K){20} KKKKK
62: KKKKK (K){20} KKKKK
63: KKKKK (K){20} KKKKK
64: KKKKK (K){20} KKKKK
65: KKKKK (K){20} KKKKK
ABG07742 ck: 8672 len: 502 ! Abg07742 Novel human diagnostic protein #77
54: DDEEE (R,K){20,20} KKKKK
55: DEBER (R,K){20} KKKKK
56: EEERR (R,K){20} KKKKK
57: EERRR (R,K){20} KKKKK
58: EERRK (R,K){20} KKKKK
59: RRRKK (R,K){20} KKKKK
60: RRRKK (R,K){20} KKKKK
61: RKKKK (R,K){20} KKKKK
62: KKKKK (R,K){20} KKKKK
63: KKKKK (R,K){20} KKKKK
64: KKKKK (R,K){20} KKKKK
65: KKKKK (R,K){20} KKKKK

1

66: KKKKK (R,K){20} KKKKK
67: KKKKK (R,K){20} KKKKK
ABG10052 ck: 7107 len: 39 ! Abg10052 Novel human diagnostic protein #1
2: Q KKKKK (R,K){20,20} KKKKK
3: QK KKKKK (K){20} KKKKK
4: QKK KKKKK (K){20} KKKKK
5: QKKK KKKKK (K){20} KKKKK
6: QKKKK KKKKK (K){20} KKKKK
7: KKKKK KKKKK (K){20} KKKKK
ABG10053 ck: 3274 len: 189 ! Abg10053 Novel human diagnostic protein #1
70: EKEKE (R,K){20,20} KKKKK
71: KEKEK (R,K){20} KKKKK
72: EKEKK RKKKK (R,K){20} KKKKK
73: KEKKR KKKKK (R,K){20} KKKKK
74: EKRRK KKKKK (R,K){20} KKKKK
75: KKKKK KKKKK (R,K){20} KKKKK
76: KKKKK RKKKK (R,K){20} KKKKK
77: RKKRR KKKKK (K){20} KKKKK
78: KKKRK KKKKK (K){20} KKKKK
79: KKKKK KKKKK (K){20} KKKKK
ABG11241 ck: 3870 len: 121 ! Abg11241 Novel human diagnostic protein #1
44: KKKKK (R,K){20,20} KKKKK
45: KKEEK KKKKK (K){20} KKKKK

1

46: KEEKK KKKKKKKKKKKKKKKKKKKKK KKKKE
(K) {20}
47: BEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
48: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

ABG11242 ck: 4493 len: 100 ! Abg11242 Novel human diagnostic protein #11

(R,K) {20,20}
42: ETPSE KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
43: TPSEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
44: PSEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
45: SEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
46: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
47: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
48: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
50: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
53: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
69: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
70: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
71: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

ABG11245 ck: 2517 len: 85 ! Abg11245 Novel human diagnostic protein #11

1

33: EEEEE (R,K) {20,20}
(K) {20}
34: EEEEEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
35: EEEEEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
36: EEEEEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
37: EEEEEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

45: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
46: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
47: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
48: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
49: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
50: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
51: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
52: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}

ABG11250 ck: 4343 len: 92 ! Abg11250 Novel human diagnostic protein #11

53: EKEKE KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20,20}
(K) {20}
54: KEKEK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
55: EKEKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
56: KEKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
57: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}

ABG11266 ck: 9563 len: 146 ! Abg11266 Novel human diagnostic protein #11

36: KEKRE KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20,20}
(K) {20}
37: EKREK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
38: KREKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
39: REKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
40: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K) {20}
79: EEEEE KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
80: EEEBK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}

81: EEEKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
82: BEKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
83: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
87: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
88: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
89: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
90: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
91: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
92: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
93: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}

ABG11277 ck: 8026 len: 1,080 ! Abg11277 Novel human diagnostic protein #1

709: QKEKE KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20,20}
(R,K) {20}
710: KEKEK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
711: EKEKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
712: KEKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
713: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
714: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
715: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
716: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
717: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}
718: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K) {20}

```

719: KKKKK (R, K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
720: KKKKK (R, K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
721: KKKKK (R, K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
722: KKKKK (R, K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
723: KKKKK (R, K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
724: KKKKK (R, K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
725: KKKKK (R, K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
726: KKKKK (R, K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK
727: KKKKK (R, K) {20} KKKKK KKKKK KKKKK KKKKK KKKKK

```

ABG11732 ck: 2886 len: 56 ! Abg11732 Novel human diagnostic protein #11

	$(R, K) \{20, 20\}$	
	$(R) \{20\}$	
8:	KRRG RRRRRRRRRRRRRR RRRR	
	$(R) \{20\}$	
9:	RRGR RRRRRRRRRRRRRR RRRR	
	$(R) \{20\}$	
10:	RRGR RRRRRRRRRRRRRR RRRR	
	$(R) \{20\}$	
11:	RRRR RRRRRRRRRRRRRR RRRR	
	$(R) \{20\}$	
12:	RRRR RRRRRRRRRRRRRR RRRR	
	$(R) \{20\}$	
13:	RRRR RRRRRRRRRRRRRR RRRR	
	$(R) \{20\}$	

ABG11734 ck: 4548 len: 142 ! Abg11734 Novel human diagnostic protein #11

[illegible]

```

116: XXXX XXXXXXXXXXXXXXXXXXXXXXXX KLYF
      (K) {20}

117: XXXX XXXXXXXXXXXXXXXXXXXXXXXX KLYQ
      (K) {20}

118: XXXX XXXXXXXXXXXXXXXXXXXXXXXX LYQT
      (K) {20}

```

ABG11738 ck: 3009 len: 567 ! Abg11738 Novel human diagnostic protein #1:

[illegible]

ABG22512 ck: 8641 len: 856 | Abg22512 Novel human diagnostic protein #22

[illegible]

ABG22638 ck: 4197 len: 896 ! Abg22638 Novel human diagnostic protein #22

(R, K) {20, 20}
(R, K) {20}
227: GREGRRRRRRRRRRRRRRRRRRR

228: RERER RRRRRRRRRRRRRRRRRR RRRGX
(R,K){20}
229: ERERR RRRRRRRRRRRRRRRRRR RRGXE
(R,K){20}
230: RERRR RRRRRRRRRRRRRRRRRR RGXEF
(R,K){20}
231: ERERR RRRRRRRRRRRRRRRRRR GXEFL
(R,K){20}

ABG26213 ck: 6773 len: 735 ! Abg26213 Novel human diagnostic protein #26

173: RGSSS KKKRRKKKKKKKKKKRRR KNRKK
(R,K){20,20}
(R,K){20}
174: GSSSK KKKRRKKKKKKKKRRR NRKKK
(R,K){20}

ABG26488 ck: 523 len: 124 ! Abg26488 Novel human diagnostic protein #26

91: EEEEE RRRRRKKKKKKKKKKKKKK RKKKK
(R,K){20,20}
(R,K){20}
92: EEEER RRRRRKKKKKKKKKKKKKK RKKKK
(R,K){20}
93: EEEER RRRRRKKKKKKKKKKKKKK RKKKK
(R,K){20}
94: EEEEE RRRRRKKKKKKKKKKKKKK RKKKK
(R,K){20}
95: EEEEE RRRRRKKKKKKKKKKKKKK RKKKK
(R,K){20}
96: RRRRR KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
97: RRRRR KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
98: RRRRR KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
99: RRRRR KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
100: RKKKK KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}

ABG26489 ck: 9923 len: 120 ! Abg26489 Novel human diagnostic protein #26

(R,K){20,20}
(R){20}

70: RRRRG RRRRRRRRRRRRRRRRRR RKKKK
(R){20}
71: RRRGR RRRRRRRRRRRRRRRRRR RKKKE
(R){20}
72: RRRGR RRRRRRRRRRRRRRRRRR RKKKE
(R,K){20}
73: RRRGR RRRRRRRRRRRRRRRRRR RKKKE
(R,K){20}
74: RRRGR RRRRRRRRRRRRRRRRRR RKKKE
(R,K){20}
75: RRRRR RRRRRRRRRRRRRRRRRR RKKKE
(R,K){20}

ABG26490 ck: 4146 len: 96 ! Abg26490 Novel human diagnostic protein #2

39: KKGEE RRRRRRRRRRRRRRRRRR RRRGR
(R,K){20,20}
(R){20}
40: KGEER RRRRRRRRRRRRRRRRRR RRRGR
(R){20}
41: GEERR RRRRRRRRRRRRRRRRRR RRRGR
(R){20}
42: EEEEE RRRRRRRRRRRRRRRRRR RRRGR
(R){20}
73: GRGRG RRRRRRRRRRRRRRRRRR RRRR
(R){20}
74: RRRGR RRRRRRRRRRRRRRRRRR RRRR
(R){20}
75: GRGRG RRRRRRRRRRRRRRRRRR RRRR
(R){20}
76: RRRGR RRRRRRRRRRRRRRRRRR RRRR
(R){20}
77: RRRRR RRRRRRRRRRRRRRRRRR RRRR
(R){20}

ABG26491 ck: 8179 len: 109 ! Abg26491 Novel human diagnostic protein #2

75: RRRGE KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20,20}
(R,K){20}
76: RRRGE KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
77: RRRGE KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
78: RRRGE KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
79: RRRGE KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
80: RRRGR KKKKKKKKKKKKKKKKKKK RKKKN
(K){20}
81: RRRGR KKKKKKKKKKKKKKKKKKK RKKKN
(K){20}

82: KRRKK KKKKKKKKKKKKKKKKKKK KGNLS
(K){20}

83: KRRKK KKKKKKKKKKKKKKKKKKK KNLVS
(K){20}

ABG26492 ck: 5234 len: 68 ! Abg26492 Novel human diagnostic protein #26

(R,K){20,20}

29: KKKEE RRRRRRRRRRRRRRRRRRR RRRRG
(R){20}

30: KKEER RRRRRRRRRRRRRRRRRRR RRRGK
(R){20}

31: KEERR RRRRRRRRRRRRRRRRRRR RRGKK
(R){20}

32: EERRR RRRRRRRRRRRRRRRRRRR RGKGD
(R){20}

33: ERRRR RRRRRRRRRRRRRRRRRRR GKDDG
(R){20}

ABG26493 ck: 4204 len: 80 ! Abg26493 Novel human diagnostic protein #26

(R,K){20,20}

(R,K){20}

27: EKEKE KRRRRRRRRRRRRRRRRRR RRRRR

28: KEKEK RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

29: EKEKR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

30: KEKRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

31: EKRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

32: KRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

33: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

34: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

35: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

36: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

37: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

38: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

39: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

40: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

41: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

42: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

43: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

44: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

45: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

46: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

47: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

48: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

49: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

50: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

51: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

52: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

53: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

ABG26496 ck: 8829 len: 90 ! Abg26496 Novel human diagnostic protein #26

(R,K){20,20}

(R,K){20}

61: EKKEE KRRRRRRRRRRRRRRRRRR RRRRR

62: KKEEK RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

63: KEEKR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

64: EEKRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

65: EKRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

66: KRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

67: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

68: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

69: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

70: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

ABG26497 ck: 9186 len: 115 ! Abg26497 Novel human diagnostic protein #26

72: RRRR RRRR(R){20} RRRR
73: RRRR RRRR(R){20} RRRR
74: RRRR RRRR(R){20} RRRR
75: RRRR RRRR(R){20} RRRR
76: RRRR RRRR(R){20} RRRR
77: RRRR RRRR(R){20} RRRR
78: RRRR RRRR(R){20} RRRR
79: RRRR RRRR(R){20} RRRR
80: RRRR RRRR(R){20} RRRR
81: RRRR RRRR(R){20} RRRR
82: RRRR RRRR(R){20} RRRR
83: RRRR RRRR(R){20} RRRR
84: RRRR RRRR(R){20} RRRR
85: RRRR RRRR(R){20} RRRR
86: RRRR RRRR(R){20} RRRR
ABG26498 ck: 7156 len: 140 ! Abg26498 Novel human diagnostic protein #26
97: RGGG RRRR(R,K){20,20} RRRR
98: RGGG RRRR(R){20} RRRR
99: GGGG RRRR(R){20} RRRR
100: GGGG RRRR(R){20} RRRR
101: GRRR RRRR(R){20} RRRR
102: RRRR RRRR(R){20} RRRR
103: RRRR RRRR(R){20} RRRR
104: RRRR RRRR(R){20} RRRR

1

105: RRRR RRRR(R){20} RRRR
106: RRRR RRRR(R){20} RRRR
107: RRRR RRRR(R){20} RRRR
108: RRRR RRRR(R){20} RRRR
109: RRRR RRRR(R){20} RRRR
110: RRRR RRRR(R){20} RRRR
111: RRRR RRRR(R){20} RRRR
112: RRRR RRRR(R){20} RRRR
113: RRRR RRRR(R,K){20} RRRR
ABG26500 ck: 7252 len: 78 ! Abg26500 Novel human diagnostic protein #26
35: GGGG RRRR(R,K){20,20} RRRR
36: GGGG RRRR(R){20} RRRR
37: GGGG RRRR(R){20} RRRR
38: GGGG RRRR(R){20} RRRR
39: GGGG RRRR(R){20} RRRR
40: GGGG RRRR(R){20} RRRR
41: GGGG RRRR(R){20} RRRR
42: GGGG RRRR(R){20} RRRR
43: GGGG RRRR(R){20} RRRR
44: GGGG RRRR(R){20} RRRR
45: GGGG RRRR(R){20} RRRR
46: GGGG RRRR(R){20} RRRR
47: GGGG RRRR(R){20} RRRR
48: GGGG RRRR(R){20} RRRR

1

ABG26501 ck: 1730 len: 182 ! Abg26501 Novel human diagnostic protein #26

1
72: KEEKE KKKKKKKKKKKKKKKKKKKKK KKKKE
(R,K){20,20}
(K){20}
73: EEKEK KKKKKKKKKKKKKKKKKKKKK KKKKE
(K){20}
74: EKEKK KKKKKKKKKKKKKKKKKKKKK KKEEE
(K){20}
75: KEKKK KKKKKKKKKKKKKKKKKKKKK KEEEE
(K){20}
76: EKKKK KKKKKKKKKKKKKKKKKKKKK EEEEE
(K){20}
133: GRRRS RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
134: RRRSR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
135: RRSRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
136: RSRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
137: SRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
138: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
139: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
140: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
141: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
142: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
143: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
144: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
145: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
146: RRRRR RRRRRRRRRRRRRRRRRRRR RRRKY
(R){20}
147: RRRRR RRRRRRRRRRRRRRRRRRRR RRRYP
(R){20}
148: RRRRR RRRRRRRRRRRRRRRRRRRR RRRYP
(R){20}
149: RRRRR RRRRRRRRRRRRRRRRRRRR RRRYP
(R){20}
150: RRRRR RRRRRRRRRRRRRRRRRRRR RRRYP
(R,K){20}

ABG26502 ck: 3399 len: 101 ! Abg26502 Novel human diagnostic protein #2:

1
59: EEEEG RRRRRKKKKKKKKKKKKKK KKKKR
(R,K){20,20}
(R,K){20}
60: EEEGR RRRRRKKKKKKKKKKKKKK KKKRK
(R,K){20}
61: EEEGR RRRRRKKKKKKKKKKKKKK KKKRK
(R,K){20}
62: EGGRR RRRRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
63: GRRRR RRRRRKKKKKKKKKKKKKK RKKKK
(R,K){20}
64: RRRRR RRRRRKKKKKKKKKKKKKK RKKKK
(R,K){20}
65: RRRRR RRRRRKKKKKKKKKKKKKK KKKKE
(R,K){20}
66: RRRRR RRRRRKKKKKKKKKKKKKK KKKKE
(R,K){20}
67: RRRRR RRRRRKKKKKKKKKKKKKK KKKKE
(R,K){20}
68: RRRRR RRRRRKKKKKKKKKKKKKK KKKKE
(R,K){20}
69: RRRRR RRRRRKKKKKKKKKKKKKK KKKKE
(R,K){20}

ABG26505 ck: 4704 len: 93 ! Abg26505 Novel human diagnostic protein #2:

1
60: KEEEG RRRRRRRRRRRRRRRRRRRR RRRRR
(R,K){20,20}
(R){20}
61: EEEGR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
62: EEEGR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
63: EGGRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
64: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
65: RRRRR RRRRRRRRRRRRRRRRRRRR RKKKE
(R){20}
66: RRRRR RRRRRRRRRRRRRRRRRRRR KKKKE
(R){20}
67: RRRRR RRRRRRRRRRRRRRRRRRRR KKKKE
(R,K){20}
68: RRRRR RRRRRRRRRRRRRRRRRRRR KKKKE
(R,K){20}
69: RRRRR RRRRRRRRRRRRRRRRRRRR KKKKE
(R,K){20}

ABG26506 ck: 3684 len: 85 ! Abg26506 Novel human diagnostic protein #2:

1

ABG26507 ck: 9838 len: 109 ! Abg26507 Novel human diagnostic protein #26

(R,K){20,20}
(R,K){20}

47: EKEKE KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

48: KEKEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

49: EKEKR KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

50: KEKKR KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

51: EKKRK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

52: KKKRK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

53: RKKRK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

54: KKKRK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

55: RKKRK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

56: KKKRK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

57: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

1

ABG26507 ck: 9838 len: 109 ! Abg26507 Novel human diagnostic protein #26

(R,K){20,20}
(R){20}

60: EEEEE RRRRRRRRRRRRRRRRRRRR RRRKK
(R){20}

61: EEEER RRRRRRRRRRRRRRRRRRRR RKKKE
(R){20}

62: EEEER RRRRRRRRRRRRRRRRRRRR RKKER
(R){20}

63: EEEEE RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}

64: EEEEE RRRRRRRRRRRRRRRRRRRR KKKKK
(R,K){20}

65: RRRRR RRRRRRRRRRRRRRRRRRRR ERKKE

1

ABG26508 ck: 4488 len: 121 ! Abg26508 Novel human diagnostic protein #26

(R,K){20,20}
(R,K){20}

56: DDEEE RRRKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

57: DEBER RRRKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

58: EEERR RKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

59: EEEER KKKKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}

60: EERRR KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

61: RRRRK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

62: RRRKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

63: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

67: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

68: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

69: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

1

ABG26510 ck: 3426 len: 74 ! Abg26510 Novel human diagnostic protein #26

(R,K){20,20}
(R){20}

47: RRRRS RRRRRRRRRRRRRRRRRRRR RRRRN
(R){20}

48: RRRSR RRRRRRRRRRRRRRRRRRRR RRRRN
(R){20}

49: RRSRR RRRRRRRRRRRRRRRRRRRR RRRRN
(R){20}

50: RSRRR RRRRRRRRRRRRRRRRRRRR RRRRN
(R){20}

51: SRRRR RRRRRRRRRRRRRRRRRRRR RRRRN
(R){20}

ABG26513 ck: 3117 len: 265 ! Abg26513 Novel human diagnostic protein #26

(R,K){20,20}
(R,K){20}

199: EEEEE KKKKKKKKKKKKKKKKKKKKK EEEKK
(K){20}

222: KEEEE KKKKKKKKKKKKKKKKKKKKK KKKKE

223: KEEK KKKKKKKKKKKKKKKKKKK RKEEE
224: EEEK KKKKKKKKKKKKKKKKKKK KEEEE
225: EEEK KKKKKKKKKKKKKKKKKKK EEEEE

ABG26514 ck: 9106 len: 218 ! Abg26514 Novel human diagnostic protein #26

182: EERG RRRRRRRRRRRRRRRRRR RRRRR
183: ERGR RRRRRRRRRRRRRRRRRR RRRRR
184: RGRR RRRRRRRRRRRRRRRRRR RRRRR
185: RGRR RRRRRRRRRRRRRRRRRR RRRRR
186: GRRR RRRRRRRRRRRRRRRRRR RRRRR
187: RRRR RRRRRRRRRRRRRRRRRR RRRR
188: RRRR RRRRRRRRRRRRRRRRRR RRRR
189: RRRR RRRRRRRRRRRRRRRRRR RRRR
190: RRRR RRRRRRRRRRRRRRRRRR RRRR
191: RRRR RRRRRRRRRRRRRRRRRR RRRR
192: RRRR RRRRRRRRRRRRRRRRRR RRRR
193: RRRR RRRRRRRRRRRRRRRRRR RRRR

ABG26515 ck: 7288 len: 389 ! Abg26515 Novel human diagnostic protein #26

259: KEES RKKKKKKKKKKKKKKKKKK KKKR
260: KEES RKKKKKKKKKKKKKKKKKK KKKR
261: EESR KKKKKKKKKKKKKKKKKKK KRRN
262: ESRK KKKKKKKKKKKKKKKKKKK RRRN
263: SRKK KKKKKKKKKKKKKKKKKKK KRRN
264: RKKK KKKKKKKKKKKKKKKKKKK RRRN

265: KKKK KKKKKKKKKKKKKKKKK NKKK

ABG26516 ck: 2295 len: 91 ! Abg26516 Novel human diagnostic protein #2

45: EEEE RRRRRRRRRRRRRRRRRR KRRK
46: EEEER RRRRRRRRRRRRRRRRRR RRRR
47: EEER RRRRRRRRRRRRRRRRRR KRRK
48: EERR RRRRRRRRRRRRRRRRRR RRRR
49: EERR RRRRRRRRRRRRRRRRRR KRRK
50: RRRR RRRRRRRRRRRRRRRRRR RRRR
51: RRRR RRRRRRRRRRRRRRRRRR KRRK
52: RRRR RRRRRRRRRRRRRRRRRR KRRK
53: RRRR RRRRRRRRRRRRRRRRRR KRRK
54: RRRR RRRRRRRRRRRRRRRRRR KRRK

ABG26518 ck: 8431 len: 761 ! Abg26518 Novel human diagnostic protein #2

725: EEEE RRRRRRRRRRRRRRRRRR RRRR
726: EEEER RRRRRRRRRRRRRRRRRR RRRR
727: EEER RRRRRRRRRRRRRRRRRR RRRR
728: EERR RRRRRRRRRRRRRRRRRR RRRR
729: EERR RRRRRRRRRRRRRRRRRR RRRR
730: RRRR RRRRRRRRRRRRRRRRRR RRRR
731: RRRR RRRRRRRRRRRRRRRRRR RRRR
732: RRRR RRRRRRRRRRRRRRRRRR RRRR
733: RRRR RRRRRRRRRRRRRRRRRR RRRR
734: RRRR RRRRRRRRRRRRRRRRRR RRRR
735: RRRR RRRRRRRRRRRRRRRRRR RRRR

```

736: RRRR RRRRRRRRRRRRRRRR KKKK
      (R, K) {20}

737: RRRR RRRRRRRRRRRRRRRR KKKK
      (R, K) {20}

738: RRRR RRRRRRRRRRRRRRRR KKK
      (R, K) {20}

739: RRRR RRRRRRRRRRRRRRRR KKK
      (R, K) {20}

740: RRRR RRRRRRRRRRRRRRRR KK
      (R, K) {20}

741: RRRR RRRRRRRRRRRRRRRR N
      (R, K) {20}

```

ABG26520 ck: 2487 len: 99 ! Abg26520 Novel human diagnostic protein #268

1

[illegible][illegible]

76: RRRR RRRRRRRRRRRRRRRR RKKI
(R){20}
77: RRRR RRRRRRRRRRRRRRRR RKKI
(R){20}
78: RRRR RRRRRRRRRRRRRRRR KI
(R,K){20}
79: RRRR RRRRRRRRRRRRRRRR I
(R,K){20}

ABG26521 ck: 482 len: 367 ! Abg26521 Novel human diagnostic protein #26

317: EEEG KKKKKKKKKKKKKKKKK KKKR
(R,K){20,20}
(K){20}
318: EEEG KKKKKKKKKKKKKKKKK KKKR
(K){20}
319: EEEG KKKKKKKKKKKKKKKKK KKKR
(K){20}
320: EEEG KKKKKKKKKKKKKKKKK KKKR
(K){20}
321: GKKK KKKKKKKKKKKKKKKKK RKKK
(R,K){20}
322: KKKK KKKKKKKKKKKKKKKKK RKKK
(R,K){20}
323: KKKK KKKKKKKKKKKKKKKKK RKKK
(R,K){20}
324: KKKK KKKKKKKKKKKKKKKKK KKKK
(R,K){20}
325: KKKK KKKKKKKKKKKKKKKKK KKEQ
(R,K){20}
326: KKKK KKKKKKKKKKKKKKKKK KKEQ
(R,K){20}
327: KKKK KKKKKKKKKKKKKKKKK KEOQ
(R,K){20}
328: KKKK KKKKKKKKKKKKKKKKK EOQL
(R,K){20}

ABG26522 ck: 2060 len: 152 ! Abg26522 Novel human diagnostic protein #26

46: EEEE RRRRRRRRRRRRRRRR RKKK
(R,K){20,20}
(R){20}
47: EEEER RRRRRRRRRRRRRRRR RKKEG
(R){20}
48: EEEER RRRRRRRRRRRRRRRR KKEGE
(R){20}
49: EEEER RRRRRRRRRRRRRRRR KEGEE
(R,K){20}
50: EEEER RRRRRRRRRRRRRRRR EGEEG
(R,K){20}
96: KKEG RRRRRRRRRRRRRRRR RRRR
(R){20}

97: KEEG RRRRRRRRRRRRRRRR RRRR
(R){20}
98: EEEG RRRRRRRRRRRRRRRR RRRR
(R){20}
99: EEEG RRRRRRRRRRRRRRRR RRRR
(R){20}
100: GRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
101: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
102: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
103: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
104: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
105: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
106: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
107: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
108: RRRR RRRRRRRRRRRRRRRR RRRK
(R){20}
109: RRRR RRRRRRRRRRRRRRRR RRRK
(R){20}
110: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
111: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
112: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
113: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}

ABG26525 ck: 5433 len: 103 ! Abg26525 Novel human diagnostic protein #2

3: RD RRRRRRRRRRRRRRRR RRRR
(R,K){20,20}
(R,K){20}

ABG26526 ck: 1887 len: 115 ! Abg26526 Novel human diagnostic protein #2

48: EEEE RRRRRRRRRRRRRRRR RRRR
(R,K){20,20}
(R){20}
49: EEEER RRRRRRRRRRRRRRRR RRRR
(R,K){20}
50: EEEER RRRRRRRRRRRRRRRR RRRR
(R,K){20}
51: EEEER RRRRRRRRRRRRRRRR RRRR
(R,K){20}

18: KKKKK (R,K){20}
 19: KKKKK (R,K){20}
 20: KKKKK (R,K){20}
 21: KKKKK (R,K){20}

ABG26717 ck: 4415 len: 78 ! Abg26717 Novel human diagnostic protein #26

37: QMLSV KKKKK (R,K){20,20}
 38: MLSVK KKKKK (R,K){20}
 39: LSVKK KKKKK (R,K){20}
 40: SVKKK KKKKK (R,K){20}
 41: VKKKK KKKKK (R,K){20}
 42: KKKKK KKKKK (R,K){20}
 43: KKKKK KKKKK (K){20}
 44: KKKKK KKKKK (K){20}
 45: KKKKK KKKKK (K){20}
 46: KKKKK KKKKK (K){20}
 47: KKKKK KKKKK (K){20}
 48: KKKKK KKKKK (K){20}
 49: KKKKK KKKKK (K){20}
 50: KKKKK KKKKK (K){20}
 51: KKKKK KKKKK (K){20}
 52: KKKKK KKKKK (K){20}
 53: KKKKK KKKKK (K){20}
 54: KKKKK KKKKK (K){20}

ABG26718 ck: 9531 len: 141 ! Abg26718 Novel human diagnostic protein #26

80: EVARP KKKKK (R,K){20,20}
 81: VARPR KKKKK (K){20}
 82: ARPRK KKKKK (K){20}
 83: RPRKK KKKKK (K){20}
 84: PRKKK KKKKK (K){20}
 85: RKKKK KKKKK (K){20}
 86: KKKKK KKKKK (K){20}
 87: KKKKK KKKKK (K){20}
 88: KKKKK KKKKK (K){20}
 89: KKKKK KKKKK (K){20}

ABG26719 ck: 72 len: 83 ! Abg26719 Novel human diagnostic protein #2

42: ETPSE KKKKK (R,K){20,20}
 43: TPSEK KKKKK (K){20}
 44: PSEKK KKKKK (K){20}
 45: SEKKK KKKKK (K){20}
 46: EKKKK KKKKK (K){20}
 47: KKKKK KKKKK (K){20}
 48: KKKKK KKKKK (K){20}
 49: KKKKK KKKKK (K){20}
 50: KKKKK KKKKK (K){20}
 51: KKKKK KKKKK (K){20}
 52: KKKKK KKKKK (K){20}
 53: KKKKK KKKKK (K){20}
 54: KKKKK KKKKK (K){20}

1

(R,K){20,20}
(K){20}
259: KEEKE KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
260: BEEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
261: BEEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
262: EKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
263: EKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
264: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
265: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
266: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
267: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
268: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
269: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
270: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
271: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
272: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
273: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
274: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
275: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
276: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

ABG26725 ck: 909 len: 131 ! Abg26725 Novel human diagnostic protein #26

1

(R,K){20,20}
(R,K){20}
70: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
71: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
72: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
73: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1

ABG26726 ck: 930 len: 677 ! Abg26726 Novel human diagnostic protein #26

74: EKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
75: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
76: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
77: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
78: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
79: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
80: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
81: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
82: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
83: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1

(R,K){20,20}
(R,K){20}
240: EKEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
241: KEKE KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
242: EKEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
243: KEKE KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
244: EKEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
245: KEKE KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
246: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
247: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
248: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
249: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
250: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
251: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
252: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

253: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
254: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
255: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
256: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
257: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
258: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
259: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
260: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R, K) {20}
261: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R, K) {20}
262: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R, K) {20}
263: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R, K) {20}
264: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R, K) {20}
265: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R, K) {20}

ABG26727 ck: 813 len: 329 ! Abg26727 Novel human diagnostic protein #26

1

215: EKEKE KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R, K) {20,20}
 (K) {20}
216: KEKEK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
217: EKEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
218: KEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
219: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
220: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
221: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
222: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
223: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
224: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

225: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
226: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
227: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKTS
 (K) {20}
228: KKKKK KKKKKKKKKKKKKKKKKKKKK KRTSQ
 (K) {20}
229: KKKKK KKKKKKKKKKKKKKKKKKKKK KTSQD
 (K) {20}
230: KKKKK KKKKKKKKKKKKKKKKKKKKK TSQDA
 (K) {20}

ABG26879 ck: 2669 len: 800 ! Abg26879 Novel human diagnostic protein #2

1

559: EREQE KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R, K) {20,20}
 (R, K) {20}
560: REQEK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R, K) {20}
561: EQEKK KKKKKKKKKKKKKKKKKKKKK KKKQQ
 (R, K) {20}
562: QEKKK KKKKKKKKKKKKKKKKKKKKK KKKQE
 (K) {20}
563: EKKKK KKKKKKKKKKKKKKKKKKKKK KQEEK
 (K) {20}
564: KKKKK KKKKKKKKKKKKKKKKKKKKK QOEKE
 (K) {20}

ABG28885 ck: 8668 len: 137 ! Abg28885 Novel human diagnostic protein #2:

1

35: EEEEG RRRRRRRRRRRRRRRRRRRRR RRRRK
 (R, K) {20,20}
 (R) {20}
36: EEEGR RRRRRRRRRRRRRRRRRRRRR RRRKK
 (R) {20}
37: EGGRR RRRRRRRRRRRRRRRRRRRRR RRRKV
 (R) {20}
38: EGRRR RRRRRRRRRRRRRRRRRRRRR RKKVL
 (R) {20}
39: GRRRR RRRRRRRRRRRRRRRRRRRRR KVVLS
 (R) {20}
40: RRRRR RRRRRRRRRRRRRRRRRRRRR KVLSP
 (R, K) {20}
41: RRRRR RRRRRRRRRRRRRRRRRRRRR VLSPS
 (R, K) {20}

AAG65985 ck: 8085 len: 154 ! Aag65985 B726P splice variant sequence. 2/:

1

114: TQLRQ KKKKKKKKKKKKKKKKKKKKK KKKKK
 (R, K) {20,20}
 (K) {20}
115: QLFRQ KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

[illegible][illegible]

1

```

1
(R,K){20,20}
(R,K){20}
23: EEEEE KKKKKKKKKKKKKKKKKKK EEEK
(K){20}
46: KKEEE KKKKKKKKKKKKKKKKKKK KKEE
(K){20}
47: KEEK KKKKKKKKKKKKKKKKKKK RKEE
(R,K){20}
48: EEEK KKKKKKKKKKKKKKKKKKK KEEE
(R,K){20}
49: EKKK KKKKKKKKKKKKKKKKKKK EEEE
ABB2645 ck: 1939 len: 130 | Abb2645 Peptide #2296 encoded by breast ca
1
(R,K){20,20}
(R,K){20}
42: EGRK RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
43: GRKR RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
44: RKRR RRRRRRRRRRRRRRRRRR RRRR

```


60: RRRR RRRRRRRRRRRRRRRRRR RNTN (R){20}
 61: RRRR RRRRRRRRRRRRRRRRRR RNTN (R){20}
 62: RRRR RRRRRRRRRRRRRRRRRR NTNNE (R){20}
 ABB33064 ck: 5383 len: 86 ! Abb33064 Peptide #570 encoded by human foet
 15: RRRG RRRRRRRRRRRRRRRRRR RRRR (R,K){20,20}
 16: RRRG RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 17: RRRG RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 18: RRRG RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 19: RRRG RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 20: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 21: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 22: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 23: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 24: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 25: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 26: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 27: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 28: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 29: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 30: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 31: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 32: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 33: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 34: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}

35: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 36: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 37: RRRR RRRRRRRRRRRRRRRRRR RRRR (R){20}
 38: RRRR RRRRRRRRRRRRRRRRRR RRRR (R){20}
 39: RRRR RRRRRRRRRRRRRRRRRR RRRR (R){20}
 40: RRRR RRRRRRRRRRRRRRRRRR RRRR (R){20}
 41: RRRR RRRRRRRRRRRRRRRRRR RRRR (R){20}
 42: RRRR RRRRRRRRRRRRRRRRRR RRRR (R){20}
 43: RRRR RRRRRRRRRRRRRRRRRR RRRR (R){20}
 44: RRRR RRRRRRRRRRRRRRRRRR RRRR (R){20}
 45: RRRR RRRRRRRRRRRRRRRRRR RRRR (R){20}
 ABB33937 ck: 1334 len: 86 ! Abb33937 Peptide #1443 encoded by human fo
 57: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20,20}
 58: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 59: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 60: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 61: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 62: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 63: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 64: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 65: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 66: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}
 67: RRRR RRRRRRRRRRRRRRRRRR RRRR (R,K){20}

1 (R,K){20,20}
(R,K){20}
33: EGRG RRRRRRRRRRRRRRRRRR RRGG
(R,K){20}
34: EGRG RRRRRRRRRRRRRRRRRR RGGG
(R,K){20}
35: GRGR RRRRRRRRRRRRRRRRRR GGRG
ABB34533 ck: 2276 len: 89 | Abb34533 Peptide #2039 encoded by human fo
(R,K){20,20}
(R,K){20}
23: EEEE KKKKKKKKKKKKKKKKKK EEEK
(K){20}
46: KEEE KKKKKKKKKKKKKKKKKK KKEE
(K){20}
47: KEEK KKKKKKKKKKKKKKKKKK KKEE
(R,K){20}
48: EEEK KKKKKKKKKKKKKKKKKK KEEE
(R,K){20}
49: EEEK KKKKKKKKKKKKKKKKKK EEEE
ABB34819 ck: 1939 len: 130 | Abb34819 Peptide #2325 encoded by human fo
(R,K){20,20}
(R,K){20}
42: EGKE RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
43: GRKR RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
44: RKRR RRRRRRRRRRRRRRRRRR RRRK
(R,K){20}
45: KRRR RRRRRRRRRRRRRRRRRR RRRK
(R,K){20}
46: ERRR RRRRRRRRRRRRRRRRRR RRRK
(R,K){20}
47: RRRR RRRRRRRRRRRRRRRRRR RRRK
(R,K){20}
48: RRRR RRRRRRRRRRRRRRRRRR RRRK
(R,K){20}
49: RRRR RRRRRRRRRRRRRRRRRR RRRK
(R,K){20}
50: RRRR RRRRRRRRRRRRRRRRRR RRRK
(R,K){20}
51: RRRR RRRRRRRRRRRRRRRRRR RRRK
(R,K){20}
52: RRRR RRRRRRRRRRRRRRRRRR RRRK
(R,K){20}
53: RRRR RRRRRRRRRRRRRRRRRR RRRK
(R,K){20}
54: RRRR RRRRRRRRRRRRRRRRRR RRRK

ABB35110 ck: 3607 len: 88 | Abb35110 Peptide #2616 encoded by human fo
(R,K){20,20}
(K){20}
39: EERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
40: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
41: KERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
42: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
43: EERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
44: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
45: KERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
46: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
47: KERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
48: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
49: EERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
50: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
51: KERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
52: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
53: EERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
54: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
55: KERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
56: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
57: KERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
58: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
59: EERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
60: RERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}
61: KERE KKKKKKKKKKKKKKKKKK KKKK
(K){20}

47: KRRR KKKKKKKKKKKKKKK KSAH
(K){20}
48: KRRR KKKKKKKKKKKKKKK KSAH
(K){20}
49: KRRR KKKKKKKKKKKKKKK SAH
(K){20}

ABB37567 ck: 1560 len: 88 ! Abb37567 Peptide #5073 encoded by human fo

1
(R,K){20,20}
43: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
44: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
45: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
46: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
47: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
48: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
49: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
50: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
51: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
52: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
53: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
54: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
55: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
56: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
57: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
58: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
59: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
60: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
61: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
62: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}

ABB37780 ck: 2324 len: 36 ! Abb37780 Peptide #5286 encoded by human fo
(R,K){20,20}
16: KRRR KRRRRRRRRRRRRRR R
(R,K){20}
17: KRRR KRRRRRRRRRRRRRR R
(R,K){20}

ABB40272 ck: 8343 len: 66 ! Abb40272 Peptide #7778 encoded by human fo

1
(R,K){20,20}
6: EERE KRRRRRRRRRRRRRR KRRR
(R,K){20}
7: EERE KRRRRRRRRRRRRRR KRRR
(R,K){20}
8: EERE KRRRRRRRRRRRRRR KRRR
(R,K){20}
9: EERE KRRRRRRRRRRRRRR KRRR
(R,K){20}

ABB42642 ck: 2394 len: 57 ! Abb42642 Peptide #10148 encoded by human fo

1
(R,K){20,20}
20: EERE KRRRRRRRRRRRRRR RRRR
(R){20}
21: EERE KRRRRRRRRRRRRRR RRRR
(R){20}
22: EERE KRRRRRRRRRRRRRR RRRR
(R){20}
23: EERE KRRRRRRRRRRRRRR RRRR
(R){20}
24: EERE KRRRRRRRRRRRRRR RRRR
(R){20}
25: EERE KRRRRRRRRRRRRRR RRRR
(R){20}

ABB43181 ck: 4228 len: 24 ! Abb43181 Peptide #10687 encoded by human fo

1
(R,K){20,20}
1: KRRR KRRRRRRRRRRRRRR RRRR
(R,K){20}
2: KRRR KRRRRRRRRRRRRRR RRRR
(R,K){20}
3: KRRR KRRRRRRRRRRRRRR RRRR
(R,K){20}
4: KRRR KRRRRRRRRRRRRRR RRRR
(R,K){20}

ABB44317 ck: 4895 len: 51 ! Abb44317 Peptide #11823 encoded by human fo

1
(R,K){20,20}
18: LRRR KRRRRRRRRRRRRRR RRRR
(R,K){20}
19: LRRR KRRRRRRRRRRRRRR RRRR
(R,K){20}

20: KPMRK RRRKKRRRRKKRRKKRRKK LTTT

ABBI17165 ck: 8887 len: 42 ! Abbi17165 Human nervous system related poly

21: FFTTE KKKKKKKKKKKKKKKKKKKKK KX
(R,K){20}

22: FTTEK KKKKKKKKKKKKKKKKKKKKK X
(K){20}

ABBI18534 ck: 5383 len: 86 ! Abbi18534 Protein #533 encoded by probe for

15: RREEG RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20,20}

16: RRRGR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}

17: RRRGR RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

18: RRRGR RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

19: GRRRR RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

20: RRRRR RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

21: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

22: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

23: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

24: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

25: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

26: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

27: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

28: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

29: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

30: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

31: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

32: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

33: RRRRK RRRRRKKKKKKKKKKKKKK RRRR
(R,K){20}

34: KKKRK KKKRRRRRRRRRRRRRRRRRR (R,K){20}

35: KKKRK KKKRRRRRRRRRRRRRRRRRR (R,K){20}

36: KKKRK KKKRRRRRRRRRRRRRRRRRR (R,K){20}

37: KKKRK KKKRRRRRRRRRRRRRRRRRR (R){20}

38: KKKRK KKKRRRRRRRRRRRRRRRRRR (R){20}

39: KKKRK KKKRRRRRRRRRRRRRRRRRR (R){20}

40: KKKRK KKKRRRRRRRRRRRRRRRRRR (R){20}

41: KKKRK KKKRRRRRRRRRRRRRRRRRR (R){20}

42: KKKRK KKKRRRRRRRRRRRRRRRRRR (R){20}

43: KKKRK KKKRRRRRRRRRRRRRRRRRR (R){20}

44: KKKRK KKKRRRRRRRRRRRRRRRRRR (R){20}

45: KKKRK KKKRRRRRRRRRRRRRRRRRR (R){20}

ABBI19373 ck: 1334 len: 86 ! Abbi19373 Protein #1372 encoded by probe fo

57: EEEEG RRRKKRRKKKKKKKKKKKK KKKK (R,K){20,20}

58: EEEGR RRRKKRRKKKKKKKKKKKK KKKK (R,K){20}

59: EEEGR RRRKKRRKKKKKKKKKKKK KKKK (R,K){20}

60: EEEGR RRRKKRRKKKKKKKKKKKK KKKK (R,K){20}

61: GRRRK KKKRRKKKKKKKKKKKKKK KKKK (R,K){20}

62: RRRRK KKKRRKKKKKKKKKKKKKK KKKK (R,K){20}

63: RRRRK KKKRRKKKKKKKKKKKKKK KKKK (R,K){20}

64: RRRRK RRRKKRRKKKKKKKKKKKK KKKK (R,K){20}

65: KKKRK KKKRRKKKKKKKKKKKKKK KK (R,K){20}

66: KKKRK RRRKKRRKKKKKKKKKKKK K (R,K){20}

67: KKKRK KKKRRKKKKKKKKKKKKKK (K){20}

1

ABBI9467 ck: 9082 len: 167 ! Abbi9467 Protein #1466 encoded by probe for
(R,K){20}
(R,K){20}

33: EBGGR RRRRRRRRRRRRRRRRRRRR RRGGG
(R,K){20}

34: EGRGR RRRRRRRRRRRRRRRRRRRR RGGGR
(R,K){20}

35: GRGRR RRRRRRRRRRRRRRRRRRRR GGGRR
(R,K){20}

ABBI9943 ck: 2276 len: 89 ! Abbi9943 Protein #1942 encoded by probe for
(R,K){20,20}
(R,K){20}

23: EEEEE KKKKKKKKKKKKKKKKKKKK EEEK
(K){20}

46: KKEEE KKKKKKKKKKKKKKKKKKKK KKEE
(K){20}

47: KEEK KKKKKKKKKKKKKKKKKKKK KKEE
(R,K){20}

48: EEEK KKKKKKKKKKKKKKKKKKKK KEEE
(R,K){20}

49: EKKK KKKKKKKKKKKKKKKKKKKK EEEE
(R,K){20}

ABB20238 ck: 1939 len: 130 ! ABB20238 Protein #2237 encoded by probe for
(R,K){20,20}
(R,K){20}

42: EGRKE RRRRRRRRRRRRRRRRRRRR RRRR
(R,K){20}

43: GRKER RRRRRRRRRRRRRRRRRRRR RRRR
(R,K){20}

44: RKERR RRRRRRRRRRRRRRRRRRRR RRRK
(R,K){20}

45: KERRR RRRRRRRRRRRRRRRRRRRR RRRK
(R,K){20}

46: ERRRR RRRRRRRRRRRRRRRRRRRR RKKK
(R,K){20}

47: RRRRR RRRRRRRRRRRRRRRRRRRR RKKK
(R,K){20}

48: RRRRR RRRRRRRRRRRRRRRRRRRR KKKK
(R,K){20}

49: RRRRR RRRRRRRRRRRRRRRRRRRR KKKK
(R,K){20}

50: RRRRR RRRRRRRRRRRRRRRRRRRR KKKK
(R,K){20}

51: RRRRR RRRRRRRRRRRRRRRRRRRR KKEE
(R,K){20}

52: RRRRR RRRRRRRRRRRRRRRRRRRR KKEE
(R,K){20}

53: RRRRR RRRRRRRRRRRRRRRRRRRR KEEE
(R,K){20}

1

54: RRRR RRRRRRRRRRRRRRRRRRRR EEEE
(R,K){20}

ABB20531 ck: 3607 len: 88 ! Abb20531 Protein #2530 encoded by probe for
(R,K){20,20}
(K){20}

39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

40: RKREK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

41: KREKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

42: REKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

46: KKKK (R,K){20} KKKKKKKKKKKKKKKKK KKKSA
47: KKKR (K){20} KKKKKKKKKKKKKKKKKKKKK KKSAA
48: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KSAH
49: KKKK (K){20} KKKKKKKKKKKKKKKKKKK SAH
ABB22862 ck: 1560 len: 88 ! Abb22862 Protein #4861 encoded by probe for
43: RRRG (R,K){20,20} RRRRRRRRRRRRRRRRR RRRRR
44: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
45: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
46: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
47: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
48: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
49: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
50: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
51: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
52: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
53: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
54: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
55: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
56: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
57: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
58: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
59: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
60: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
61: RRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR

62: RRRR (R){20} RRRRRRRRRRRRRRRRR NTNNE
ABB23064 ck: 2324 len: 36 ! Abb23064 Protein #5063 encoded by probe for
16: KERK (R,K){20,20} KKKKKKKKKKKKKKKKKKK R
17: ERKTK (R,K){20} KKKKKKKKKKKKKKKKKKK
ABB24685 ck: 8343 len: 66 ! Abb24685 Protein #6684 encoded by probe for
6: ETERE (R,K){20,20} KKKKKKKKKKKKKKKKK KKKKK
7: TEREK (R,K){20} KKKKKKKKKKKKKKKKK KKKKK
8: EREKK (R,K){20} KKKKKKKKKKKKKKKKK KKKKK
9: REKKK (R,K){20} KKKKKKKKKKKKKKKKK KKKKK
ABB25988 ck: 2394 len: 57 ! Abb25988 Protein #7987 encoded by probe for
20: EEEG (R,K){20,20} RRRRRRRRRRRRRRRRR RRRRR
21: EEEGR (R){20} RRRRRRRRRRRRRRRRR RRRRG
22: EEGRR (R){20} RRRRRRRRRRRRRRRRR RRRGR
23: EGRRR (R){20} RRRRRRRRRRRRRRRRR RRRGR
24: GRRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
25: RRRRR (R){20} RRRRRRRRRRRRRRRRR RRRRR
ABB27176 ck: 4895 len: 51 ! Abb27176 Protein #9175 encoded by probe for
18: LFKPM (R,K){20,20} RRRKKKKKKKKKKKKKK KKLTT
19: FKPMR (R,K){20} KRRKKKKKKKKKKKK KLT TT
20: KPMRK (R,K){20} RRRKKKKKKKKKKKK LTTT
ABB10296 ck: 3983 len: 292 ! Abb10296 Human cDNA SEQ ID NO: 604. 1/2002
273: QVFAP (R,K){20,20} RRRKKKKKKKKKKKKKK
ABB10485 ck: 7611 len: 315 ! Abb10485 Human cDNA SEQ ID NO: 793. 1/2002

```

1
  (R,K){20,20}
  (R,K){20}
273: QVFAP RKKKKKKKKKKKKKKKKKKKK KGRRS
274: VPAPR KKKKKKKKKKKKKKKKKKKK GGRSR

AAU21948 ck: 444 len: 66 ! Auu21948 Human cardiovascular system antigen
  (R,K){20,20}
  (K){20}
35: SMTFS KKKKKKKKKKKKKKKKKKKK KKGKK
36: MTFPSK KKKKKKKKKKKKKKKKKKKK XGKKK

AAU22148 ck: 4665 len: 34 ! Auu22148 Human cardiovascular system antigen
  (R,K){20,20}
  (K){20}
10: PELLK KKKKKKKKKKKKKKKKKKKK KKKKK
11: ELLLK KKKKKKKKKKKKKKKKKKKK KKKK
12: LLLKK KKKKKKKKKKKKKKKKKKKK KKK
13: LLKKK KKKKKKKKKKKKKKKKKKKK KK
14: LKKKK KKKKKKKKKKKKKKKKKKKK K
15: KKKKK KKKKKKKKKKKKKKKKKKKK

AAU22186 ck: 269 len: 76 ! Auu22186 Human cardiovascular system antigen
  (R,K){20,20}
  (K){20}
37: TPSRA KKKKKKKKKKKKKKKKKKKK KKKKK
38: PSRAK KKKKKKKKKKKKKKKKKKKK KKKKK
39: SRAKK KKKKKKKKKKKKKKKKKKKK KKKKK
40: RAKKK KKKKKKKKKKKKKKKKKKKK KKKKK
41: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKI
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKIK

AAU22374 ck: 8278 len: 53 ! Auu22374 Human cardiovascular system antigen
  (R,K){20,20}
  (K){20}

```

```

30: NCGIL KKKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
31: CGILK KKKKKKKKKKKKKKKKKKKK KKK
  (K){20}
32: GILKK KKKKKKKKKKKKKKKKKKKK KK
  (K){20}
33: ILKKK KKKKKKKKKKKKKKKKKKKK K
  (K){20}
34: LKKKK KKKKKKKKKKKKKKKKKKKK

AAU23799 ck: 6158 len: 272 ! Auu23799 Novel human enzyme polypeptide #8
  (R,K){20,20}
  (K){20}
238: SPANA KKKKKKKKKKKKKKKKKKKK KKKKG
  (K){20}
239: FANAK KKKKKKKKKKKKKKKKKKKK KKKGR
  (K){20}
240: ANAKK KKKKKKKKKKKKKKKKKKKK KGRGP
  (K){20}
241: NAKKK KKKKKKKKKKKKKKKKKKKK KGRPX
  (K){20}
242: AKKKK KKKKKKKKKKKKKKKKKKKK GRPXX

AAU27944 ck: 1121 len: 69 ! Auu27944 Human contig polypeptide sequence
  (R,K){20,20}
  (R,K){20}
40: VPPLT RKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
41: PPLTR KKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
42: PLTRK KKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
43: LTRKK KKKKKKKKKKKKKKKKKKKK KKKRG
  (K){20}
44: TRKKK KKKKKKKKKKKKKKKKKKKK KKGGA
  (K){20}
45: RKKKK KKKKKKKKKKKKKKKKKKKK KRGAL
  (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK RGAL
  (R,K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK GAL

AAU31467 ck: 4264 len: 657 ! Auu31467 Novel human secreted protein #195:
  (R,K){20,20}
  (R){20}
19: RRRRP RRRRRRRRRRRRRRRRRRR RRRRL
  (R){20}
20: RRRPR RRRRRRRRRRRRRRRRRRR RRRLG
  (R){20}
21: RRRPR RRRRRRRRRRRRRRRRRRR RRLGL

```

(R){20}
22: RPRRR RRRRRRRRRRRRRRRR RLGLE

(R){20}
23: PRRRR RRRRRRRRRRRRRRRRRR LGLER

AAU33348 ck: 8085 len: 154 | Aau33348 Human breast cancer protein encode
(R,K){20,20}
(K){20}

114: TOLRQ KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
115: QLRQK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
116: LFRQK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
117: RQKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
118: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
121: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
122: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
123: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
124: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
125: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
126: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
127: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
128: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AAE09664 ck: 1663 len: 87 | Aae09664 Human pancreatic related protein H
(R,K){20,20}
(K){20}

36: KWSSX KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
37: WSSXX KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
38: SSXXX KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
39: SXXXX KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

40: XKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKRG

(K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKGX

(K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKK KRGXP

(K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKK RGXPP

(R,K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKK GXPPX

1	111: IHLNL KKKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20} (K){20}	76: LGGQK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
	112: HLNLK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}	77: GGQKK KKKKKKKKKKKKKKKKKKKKK KKKKG (K){20}
	113: LNLKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}	78: GQKKK KKKKKKKKKKKKKKKKKKKKK KKKGX (K){20}
	114: NLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}	79: QKKKK KKKKKKKKKKKKKKKKKKKKK KKGXP (K){20}
	AAM96607 ck: 4751 len: 80 ! Aam96607 Human reproductive system related	80: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGPF (K){20}
1	61: KKKFD KKKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20} (K){20}	AAU18171 ck: 9398 len: 48 ! Aau18171 Novel human DNA-binding protein #
	AAU18162 ck: 7907 len: 39 ! Aau18162 Novel human DNA-binding protein #9	2: Q KKKKKKKKKKKKKKKKKKKKK KKGGR (R,K){20,20} (K){20}
1	9: YFEDL KKKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20} (K){20}	3: QK KKKKKKKKKKKKKKKKKKKKK KGGRX (K){20}
	10: FEDLK KKKKKKKKKKKKKKKKKKKKK KKKKG (K){20}	4: QKK KKKKKKKKKKKKKKKKKKKKK GGRXR (K){20}
	11: EDLKK KKKKKKKKKKKKKKKKKKKKK KKKGX (K){20}	AAU18177 ck: 8278 len: 53 ! Aau18177 Novel human DNA-binding protein #
	12: DLKKK KKKKKKKKKKKKKKKKKKKKK KGGXA (K){20}	30: NCGIL KKKKKKKKKKKKKKKKKKKKK KKKK (R,K){20,20} (K){20}
	13: LKKKK KKKKKKKKKKKKKKKKKKKKK KGYAA (K){20}	31: CGILK KKKKKKKKKKKKKKKKKKKKK KKK (K){20}
	14: KKKKK KKKKKKKKKKKKKKKKKKKKK GXAAL (K){20}	32: GILKK KKKKKKKKKKKKKKKKKKKKK KK (K){20}
	AAU18167 ck: 9194 len: 87 ! Aau18167 Novel human DNA-binding protein #1	33: ILKKK KKKKKKKKKKKKKKKKKKKKK K (K){20}
1	52: KIILL KKKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20} (K){20}	34: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
	53: IILLK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}	AAU18178 ck: 444 len: 66 ! Aau18178 Novel human DNA-binding protein #
	54: ILLKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}	35: SMTFS KKKKKKKKKKKKKKKKKKKKK KKGKK (R,K){20,20} (K){20}
	55: LLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}	36: MTFSK KKKKKKKKKKKKKKKKKKKKK KGGKK (K){20}
	56: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}	AAU18179 ck: 5503 len: 50 ! Aau18179 Novel human DNA-binding protein #
	57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}	30: IICLL KKKKKKKKKKKKKKKKKKKKK K (R,K){20,20} (K){20}
	AAU18168 ck: 8659 len: 104 ! Aau18168 Novel human DNA-binding protein #1	31: ICLLK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
1	75: PLGGQ KKKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20} (K){20}	AAU18184 ck: 5691 len: 108 ! Aau18184 Novel human DNA-binding protein #
		78: VRPCL KKKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20} (K){20}

79: RPCLK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
80: PCLKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
81: CLKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
87: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
88: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
89: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18192 ck: 6029 len: 63 ! Aau18192 Novel human DNA-binding protein #3

1

40: KLTLL KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20} ISWG (K) {20}

AAU18200 ck: 7170 len: 63 ! Aau18200 Novel human DNA-binding protein #4

1

37: TPSRA KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20} KKKKK (K) {20}
38: PSRAK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
39: SRAKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
40: RAKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKK
41: AKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKK
42: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KK
43: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} K
44: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18204 ck: 6110 len: 61 ! Aau18204 Novel human DNA-binding protein #5

1

(R,K) {20,20}
(K) {20}

28: RPTRP KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
29: PTRPK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
30: TRPKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
31: RPKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
32: PPKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
33: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
34: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
35: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
37: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK

AAU18205 ck: 5764 len: 74 ! Aau18205 Novel human DNA-binding protein #5

1

(R,K) {20,20}
(K) {20}

40: EFLSA KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20} KKKKK
41: FLSAK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
42: LSAKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
43: SAKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
44: AKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
45: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
46: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
47: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
48: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
49: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
50: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
51: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK
52: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20} KKKKK

53: KKKKK KKKKKKKKKKKKKKKKKKK KX
 (K) {20}
 54: KKKKK KKKKKKKKKKKKKKKKKKK X
 AAU18206 ck: 9217 len: 68 ! Aau18206 Novel human DNA-binding protein #5
 (R,K) {20,20}
 (K) {20}
 38: FLFPE KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 39: LFPEK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 40: FPEKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 41: PEKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 42: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

1

AAU18208 ck: 8162 len: 79 ! Aau18208 Novel human DNA-binding protein #5
 (R,K) {20,20}
 (R,K) {20}
 41: VRPRV RKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 42: RPRVR KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 43: PRVRK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 44: RVRKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 45: VRKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

1

AAU18237 ck: 285 len: 118 ! Aau18237 Novel human DNA-binding protein #8
 (R,K) {20,20}
 (R,K) {20}
 98: EKHQK KKKKKKKKKKKKKKKKKKK G

1

AAU18238 ck: 5509 len: 58 ! Aau18238 Novel human DNA-binding protein #8
 (R,K) {20,20}
 (K) {20}
 36: FYFVC KKKKKKKKKKKKKKKKKKK KKK
 (K) {20}
 37: YFVCK KKKKKKKKKKKKKKKKKKK KK
 (K) {20}
 38: FVCKK KKKKKKKKKKKKKKKKKKK K
 (K) {20}

1

39: VCKKK KKKKKKKKKKKKKKKKKKK
 AAU18239 ck: 9074 len: 66 ! Aau18239 Novel human DNA-binding protein #
 (R,K) {20,20}
 (K) {20}
 40: LVQCE KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 41: VQCEK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 42: QCEKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 43: CEKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 44: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

1

AAU18240 ck: 8528 len: 150 ! Aau18240 Novel human DNA-binding protein #
 (R,K) {20,20}
 (K) {20}
 113: SRNTV KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 114: RNTVK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 115: NTVKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 116: TVKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 117: VKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 118: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 119: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 120: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 121: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 122: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 123: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

1

AAU18241 ck: 7676 len: 156 ! Aau18241 Novel human DNA-binding protein #
 (R,K) {20,20}
 (K) {20}
 108: KTTWI KKKKKKKKKKKKKKKKKKK KKKKK

1

109: TTWIK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
110: TWIKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
111: WIKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
112: IKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
113: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
114: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
115: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20}

AAU18242 ck: 1736 len: 40 ! Aau18242 Novel human DNA-binding protein #8

18: LPSLK KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
19: PGS�K KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
20: GSLKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
21: SLKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18244 ck: 1109 len: 98 ! Aau18244 Novel human DNA-binding protein #9

53: QTENT KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
54: TQNTK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
55: KQTKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
56: NTKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
57: TKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
60: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
61: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

63: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18246 ck: 8102 len: 111 ! Aau18246 Novel human DNA-binding protein #

78: EFHIL KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
79: FHILK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
80: HILKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
81: ILKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
82: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
83: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
84: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
85: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
86: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
87: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
88: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
89: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
90: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
91: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
92: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAU18247 ck: 8102 len: 111 ! Aau18247 Novel human DNA-binding protein #9

78: EFHIL KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}

1

(R,K){20,20}
(K){20}
31: KQLLL KKKKKKKKKKKKKKKKKKKKK KXKXG
(K){20}
32: QLLLK KKKKKKKKKKKKKKKKKKKKK KXKGG
(K){20}
33: LLLKK KKKKKKKKKKKKKKKKKKKKK KXGGF
(K){20}
34: LLKKK KKKKKKKKKKKKKKKKKKKKK XGGF

AAU18265 ck: 3679 len: 37 ! Aau18265 Novel human DNA-binding protein #1

1

(R,K){20,20}
(K){20}
15: ISPLT KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
16: SPLTK KKKKKKKKKKKKKKKKKKKKK KX
(K){20}
17: PLTKK KKKKKKKKKKKKKKKKKKKKK X

AAU18266 ck: 657 len: 196 ! Aau18266 Novel human DNA-binding protein #1

1

(R,K){20,20}
(K){20}
169: FVXPE KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
170: VXPEK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
171: XPEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
172: FEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
173: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
174: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
175: KKKKK KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
176: KKKKK KKKKKKKKKKKKKKKKKKKKK X

AAU18267 ck: 4672 len: 57 ! Aau18267 Novel human DNA-binding protein #1

1

(R,K){20,20}
(K){20}
28: DKTFH KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: KTFHK KKKKKKKKKKKKKKKKKKKKK KKKXP
(K){20}
30: TFHKK KKKKKKKKKKKKKKKKKKKKK KXKPG
(K){20}
31: FHKKK KKKKKKKKKKKKKKKKKKKKK KXPGG
(K){20}
32: HKKKK KKKKKKKKKKKKKKKKKKKKK XPGGG

1

AAU18268 ck: 9656 len: 66 ! Aau18268 Novel human DNA-binding protein #1
(R,K){20,20}
(K){20}
38: MVISV KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
39: VISVK KKKKKKKKKKKKKKKKKKKKK KKKKE
(K){20}
40: ISVVK KKKKKKKKKKKKKKKKKKKKK KREK
(K){20}
41: SVKKK KKKKKKKKKKKKKKKKKKKKK KREKX
(K){20}
42: VKKKK KKKKKKKKKKKKKKKKKKKKK REKXX
(R,K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK EKXX

AAU18270 ck: 4665 len: 34 ! Aau18270 Novel human DNA-binding protein #1

1

(R,K){20,20}
(K){20}
10: PELLK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: ELLLK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
12: LLLKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
13: LLKKK KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
14: LKKKK KKKKKKKKKKKKKKKKKKKKK K
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK

AAU18271 ck: 7810 len: 64 ! Aau18271 Novel human DNA-binding protein #1

1

(R,K){20,20}
(K){20}
37: LKVPW KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
38: KYFWK KKKKKKKKKKKKKKKKKKKKK KKKKG
(K){20}
39: YFWKK KKKKKKKKKKKKKKKKKKKKK KKKGX
(K){20}
40: FWKKK KKKKKKKKKKKKKKKKKKKKK KXGXP
(K){20}
41: WKKKK KKKKKKKKKKKKKKKKKKKKK XGXP

AAU18272 ck: 269 len: 76 ! Aau18272 Novel human DNA-binding protein #1

1

(R,K){20,20}
(K){20}
37: TPSRA KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
38: PSRAK KKKKKKKKKKKKKKKKKKKKK KKKKK


```

39: SRANK KKKK{20} KKKK
40: RANKE KK{20} KKKK
41: AKNNE KKKK{20} KKKK
42: KNNEE KKKK{20} KKKK
43: KNNEE KKKK{20} KKKK
44: KKKKE KKKK{20} KKKK
```

```
AAU18274 ck: 1663 len: 87      ! Aau
                                     (R,K){20,20}
36: KWSXX XXXXXXXXXXXXXXXXXX XXXXK XXXXK
                                     (K){20}
37: WSSXX XXXXXXXXXXXXXXXXXX XXXXK XXXXK
                                     (K){20}
38: SSXXX XXXXXXXXXXXXXXXXXX XXXXK XXXXK
                                     (K){20}
39: SXKKK XXXXXXXXXXXXXXXXXX XXXXK XXXXK
                                     (K){20}
40: XKKKK XXXXXXXXXXXXXXXXXX XXXXK XXXXK
                                     (K){20}
41: KKKKK XXXXXXXXXXXXXXXXXX XXXXK XXXXK
                                     (K){20}
```

[illegible]

7

54: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR EEEE

AAM55931 ck: 3607 len: 88 ! Aam55931 Human brain expressed single exon

39: ERRE (R,K){20,20} (K){20} RRRRRRRRRRRRRRRRRR

40: RRRR (K){20} RRRRRRRRRRRRRRRRRR

41: RRRR (K){20} RRRRRRRRRRRRRRRRRR

42: RRRR (K){20} RRRRRRRRRRRRRRRRRR

43: RRRR (K){20} RRRRRRRRRRRRRRRRRR

44: RRRR (K){20} RRRRRRRRRRRRRRRRRR

45: RRRR (K){20} RRRRRRRRRRRRRRRRRR

46: RRRR (K){20} RRRRRRRRRRRRRRRRRR

47: RRRR (K){20} RRRRRRRRRRRRRRRRRR

48: RRRR (K){20} RRRRRRRRRRRRRRRRRR

49: RRRR (K){20} RRRRRRRRRRRRRRRRRR

50: RRRR (K){20} RRRRRRRRRRRRRRRRRR

51: RRRR (K){20} RRRRRRRRRRRRRRRRRR

52: RRRR (K){20} RRRRRRRRRRRRRRRRRR

53: RRRR (K){20} RRRRRRRRRRRRRRRRRR

54: RRRR (K){20} RRRRRRRRRRRRRRRRRR

55: RRRR (K){20} RRRRRRRRRRRRRRRRRR

56: RRRR (K){20} RRRRRRRRRRRRRRRRRR

57: RRRR (K){20} RRRRRRRRRRRRRRRRRR

1

61: RRRR (K){20} RRRRRRRRRRRRRRRRRR

62: RRRR (K){20} RRRRRRRRRRRRRRRRRR

63: RRRR (K){20} RRRRRRRRRRRRRRRRRR

64: RRRR (K){20} RRRRRRRRRRRRRRRRRR

65: RRRR (K){20} RRRRRRRRRRRRRRRRRR

66: RRRR (K){20} RRRRRRRRRRRRRRRRRR

67: RRRR (K){20} RRRRRRRRRRRRRRRRRR

68: RRRR (K){20} RRRRRRRRRRRRRRRRRR

69: RRRR (K){20} RRRRRRRRRRRRRRRRRR

AAM56490 ck: 3937 len: 85 ! Aam56490 Human brain expressed single exon

1: RRRR (R,K){20,20} (K){20} RRRRRRRRRRRRRRRRRR

2: RRRR (K){20} RRRRRRRRRRRRRRRRRR

3: RRRR (K){20} RRRRRRRRRRRRRRRRRR

4: RRRR (K){20} RRRRRRRRRRRRRRRRRR

5: RRRR (K){20} RRRRRRRRRRRRRRRRRR

6: RRRR (K){20} RRRRRRRRRRRRRRRRRR

7: RRRR (K){20} RRRRRRRRRRRRRRRRRR

8: RRRR (K){20} RRRRRRRRRRRRRRRRRR

9: RRRR (K){20} RRRRRRRRRRRRRRRRRR

10: RRRR (K){20} RRRRRRRRRRRRRRRRRR

11: RRRR (K){20} RRRRRRRRRRRRRRRRRR

AAM58224 ck: 1560 len: 88 ! Aam58224 Human brain expressed single exon

43: RRRR (R,K){20,20} (R){20} RRRRRRRRRRRRRRRRRR

43: RRRR (R){20} RRRRRRRRRRRRRRRRRR

1

44: EERGR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
45: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
46: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
47: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
48: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
49: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
50: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
51: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
52: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
53: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
54: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
55: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
56: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
57: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
58: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
59: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
60: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
61: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
62: RRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}

AAM58395 ck: 2324 len: 36 ! Aam58395 Human brain expressed single exon

1

16: KERKT KKKKKRRRRRRRRRRRRR R
 (R,K){20,20}
 (R,K){20}
17: ERKT KKKRRRRRRRRRRRRRRR
 (R,K){20}

AAM61063 ck: 8343 len: 66 ! Aam61063 Human brain expressed single exon

1

6: ETERE KKKKKKKKKKKKKKKKK KKKK
 (R,K){20,20}
 (R,K){20}

7: TEREK KKKKKKKKKKKKKKKKK KKKK
 (R,K){20}
8: EREKK KKKKKKKKKKKKKKKKK KKKK
 (R,K){20}
9: REKKK KKKKKKKKKKKKKKKKK KKKK
 (R,K){20}

AAM63533 ck: 2394 len: 57 ! Aam63533 Human brain expressed single exon

1

20: EELEG RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20,20}
 (R,K){20}
21: EEEGR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
22: EEEGR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
23: EEEGR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
24: GRRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
25: RRRRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}

AAM64090 ck: 4228 len: 24 ! Aam64090 Human brain expressed single exon

1

1: RRRRRRRRRRRRRRRRRRRR RRRR
 (R,K){20,20}
 (R,K){20}
2: R RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
3: RR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}
4: RRR RRRRRRRRRRRRRRRRRR RRRR
 (R,K){20}

AAM64863 ck: 3301 len: 52 ! Aam64863 Human brain expressed single exon

1

12: KKKKK KKKKKKKKKKKKKKKKK KKKK
 (R,K){20,20}
 (R,K){20}
13: KKKKK KKKKKKKKKKKKKKKKK KKKK
 (R,K){20}
14: KKKKK KKKKKKKKKKKKKKKKK KKKK
 (R,K){20}
15: KKKKK KKKKKKKKKKKKKKKKK KKKK
 (R,K){20}
16: KKKKK KKKKKKKKKKKKKKKKK KKKK
 (R,K){20}
17: KKKKK KKKKKKKKKKKKKKKKK KKKK
 (R,K){20}
18: KKKKK KKKKKKKKKKKKKKKKK KKKK
 (R,K){20}

19: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 20: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 21: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 22: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 23: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 24: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 25: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 26: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 27: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 28: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (K){20}
 29: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (K){20}
 30: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (K){20}
 31: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (K){20}
 AAM65355 ck: 4895 len: 51 ! Aam65355 Human brain expressed single exon
 (R,K){20,20}
 (R,K){20}
 18: LFKPM KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 19: FKPMR KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 20: KPMRK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 AAM66249 ck: 5383 len: 86 ! Aam66249 Human bone marrow expressed probe
 (R,K){20,20}
 (R,K){20}
 15: RRRRG RRRKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 16: RRRGR RRRKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 17: RRRGR RRRKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 18: RRRGR RRRKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 19: GRRRR RRRKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 20: RRRRR RRRKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}

21: RRRRK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 22: RRRRK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 23: RRRRK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 24: RRRRK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 25: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 26: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 27: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 28: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 31: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 32: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 33: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 34: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 35: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
 (R,K){20}
 37: RKKKK RRRRRRRRRRRRRRRRRRRRRRRRR
 (R){20}
 38: KKKKK RRRRRRRRRRRRRRRRRRRRRRRRR
 (R){20}
 39: KKKKK RRRRRRRRRRRRRRRRRRRRRRRRR
 (R){20}
 40: KKKKK RRRRRRRRRRRRRRRRRRRRRRRRR
 (R){20}
 41: KKKKK RRRRRRRRRRRRRRRRRRRRRRRRR
 (R){20}
 42: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R){20}
 43: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R){20}
 44: RRRRR RRRRRRRRRRRRRRRRRRRRRRRRR
 (R){20}

```

45: RRRR RRRRRRRRRRRRRR NKQTK
AAM67100 ck: 1334 len: 86 ! Aam67100 Human bone marrow expressed probe
1
57: BEESG RRRKKRKRKKKKKKKKK KKKKK
(R,K){20,20}
(R,K){20}
58: BEESG RRRKKRKRKKKKKKKKK KKKKK
(R,K){20}
59: BEESG RRRKKRKRKKKKKKKKK KKKKK
(R,K){20}
60: EGRRR KKKKKRKRKKKKKKKKK KKKKK
(R,K){20}
61: GRREK KKKKKRKRKKKKKKKKK KKKKK
(R,K){20}
62: RRRKK KKKKKRKRKKKKKKKKK KKKKK
(R,K){20}
63: RRRKK KKKKKRKRKKKKKKKKK KKKK
(R,K){20}
64: RKKKK RRRKKRKRKKKKKKKKK KK
(R,K){20}
65: KKKKK RRRKKRKRKKKKKKKKK KK
(R,K){20}
66: KKKKK RRRKKRKRKKKKKKKKK K
(R,K){20}
67: KKKKK RRRKKRKRKKKKKKKKK

AAM67180 ck: 9082 len: 167 ! Aam67180 Human bone marrow expressed probe
1
33: BEESG RRRKKRKRKKKKKKKKK RRGGG
(R,K){20,20}
(R,K){20}
34: EGRGR RRRKKRKRKKKKKKKKK RGGGR
(R,K){20}
35: GRGRR RRRKKRKRKKKKKKKKK GGGRR
(R,K){20}

AAM67717 ck: 2276 len: 89 ! Aam67717 Human bone marrow expressed probe
1
23: EEEEE KKKKKRKRKKKKKKKKK EEEKK
(R,K){20,20}
(R,K){20}
46: KKEEE KKKKKRKRKKKKKKKKK KRKEE
(K){20}
47: KEESK KKKKKRKRKKKKKKKKK RKEEE
(K){20}
48: EEEKK KKKKKRKRKKKKKKKKK KESEE
(R,K){20}
49: EEKKK KKKKKRKRKKKKKKKKK EEEEE
(R,K){20}

AAM68007 ck: 1939 len: 130 ! Aam68007 Human bone marrow expressed probe

```

```

1-
(R,K){20,20}
(R,K){20}
42: EGRKE RRRRRRRRRRRRRRRR RRRRR
(R,K){20}
43: GRKER RRRRRRRRRRRRRRRR RRRRR
(R,K){20}
44: RKERR RRRRRRRRRRRRRRRR RRRRK
(R,K){20}
45: KEERR RRRRRRRRRRRRRRRR RRRKK
(R,K){20}
46: EERRR RRRRRRRRRRRRRRRR RRRKK
(R,K){20}
47: RREER RRRRRRRRRRRRRRRR RKKKK
(R,K){20}
48: RREER RRRRRRRRRRRRRRRR KKKKK
(R,K){20}
49: RREER RRRRRRRRRRRRRRRR KKKKK
(R,K){20}
50: RRRRR RRRRRRRRRRRRRRRR KKKKE
(R,K){20}
51: RRRRR RRRRRRRRRRRRRRRR KKKEE
(R,K){20}
52: RREER RRRRRRRRRRRRRRRR KKEEE
(R,K){20}
53: RREER RRRRRRRRRRRRRRRR KESEE
(R,K){20}
54: RREER RRRRRRRRRRRRRRRR EEEEE
(R,K){20}

AAM68298 ck: 3607 len: 88 ! Aam68298 Human bone marrow expressed probe
1
(R,K){20,20}
(K){20}
39: ERKRE KKKKKKKKKKKKKKKK KKKKK
(K){20}
40: RKREK KKKKKKKKKKKKKKKK KKKKK
(K){20}
41: KREKK KKKKKKKKKKKKKKKK KKKKK
(K){20}
42: REKKK KKKKKKKKKKKKKKKK KKKKK
(K){20}
43: EKKKK KKKKKKKKKKKKKKKK KKKKK
(K){20}
44: KKKKK KKKKKKKKKKKKKKKK KKKKK
(K){20}
45: KKKKK KKKKKKKKKKKKKKKK KKKKK
(K){20}
46: KKKKK KKKKKKKKKKKKKKKK KKKKK
(K){20}
47: KKKKK KKKKKKKKKKKKKKKK KKKKK
(K){20}

```


33: KKKK KKKKKKKKKKKKKKKKK KKKK
34: KKKK (R,K){20} KKKK
35: KKKK (R,K){20} KKKK
36: KKKK (R,K){20} KKKK
37: KKKK (R,K){20} KKKK
38: KKKK (R,K){20} KKKK
39: KKKK (R,K){20} KKKK
40: KKKK (R,K){20} KKKK
41: KKKK (R,K){20} KKKK
42: KKKK (R,K){20} KKKK
43: KKKK (R,K){20} KKKK
44: KKKK (R,K){20} KKKK
45: KKKK (R,K){20} KKKK
46: KKKK (R,K){20} KKKK
47: KKKK (K){20} KKKK
48: KKKK (K){20} KKKK
49: KKKK (K){20} KKKK

AAM70678 ck: 1560 len: 88 ! Aam70678 Human bone marrow expressed probe
43: RRRG (R,K){20,20} RRRR
44: RRRG (R){20} RRRR
45: RRRG (R){20} RRRR
46: RRRR (R){20} RRRR
47: RRRR (R){20} RRRR
48: RRRR (R){20} RRRR

49: RRRR RRRRRRRRRRRRRRRR RRRR
50: RRRR (R){20} RRRR
51: RRRR (R){20} RRRR
52: RRRR (R){20} RRRR
53: RRRR (R){20} RRRR
54: RRRR (R){20} RRRR
55: RRRR (R){20} RRRR
56: RRRR (R){20} RRRR
57: RRRR (R){20} RRRR
58: RRRR (R){20} RRRR
59: RRRR (R){20} RRRR
60: RRRR (R){20} RRRR
61: RRRR (R){20} RRRR
62: RRRR (R){20} RRRR
AAM70881 ck: 2324 len: 36 ! Aam70881 Human bone marrow expressed probe
16: KERKT (R,K){20,20} R
17: ERKTK (R,K){20} RRRR

AAM73767 ck: 8343 len: 66 ! Aam73767 Human bone marrow expressed probe
6: ETESE (R,K){20,20} KKKK
7: TEREK (R,K){20} KKKK
8: EREK (R,K){20} KKKK
9: REKK (R,K){20} KKKK
AAM76347 ck: 2394 len: 57 ! Aam76347 Human bone marrow expressed probe
20: EEESG (R,K){20,20} RRRR

21: EEEGR RRRRRRRRRRRRRRRRRR RRRGR (R){20}
22: EEEGR RRRRRRRRRRRRRRRRRR RRRGR (R){20}
23: EEEGR RRRRRRRRRRRRRRRRRR RRGRR (R){20}
24: GRRRR RRRRRRRRRRRRRRRRRR RGRRR (R){20}
25: RRRRR RRRRRRRRRRRRRRRRRR GRRRR (R){20}

AAM76911 ck: 4228 len: 24 ! Aam76911 Human bone marrow expressed probe

1: RRRRRKKKKRRRRRRRRRR RKRTT (R,K){20,20}
2: R RRRRRKKKKRRRRRRRRR KRT (R,K){20}
3: RR RRRRRKKKKRRRRRRRRR RT (R,K){20}
4: RRR RRRRRKKKKRRRRRRRRR T (R,K){20}

AAM78048 ck: 4895 len: 51 ! Aam78048 Human bone marrow expressed probe

18: LFKPM RRRRRKKKKRRRRRRR KKLTT (R,K){20,20}
19: FKPMR RRRRRKKKKRRRRRRR KLTTT (R,K){20}
20: KPMRK RRRRRKKKKRRRRRRR LTTTT (R,K){20}

AAM82533 ck: 1736 len: 40 ! Aam82533 Human immune/haematopoietic antigen

18: LPSGL KKKKKKKKKKKKKKKKK KKK (R,K){20,20}
19: PGSLL KKKKKKKKKKKKKKKKK KK (K){20}
20: GSKKK KKKKKKKKKKKKKKKKK K (K){20}
21: SLKKK KKKKKKKKKKKKKKKKKK (K){20}

AAM85748 ck: 7503 len: 74 ! Aam85748 Human immune/haematopoietic antigen

45: SHLTD KKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20}
46: HLTDK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
47: LTDKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}

48: TDKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
49: DKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKK KK (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKK K (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKK (K){20}

AAM90546 ck: 6676 len: 74 ! Aam90546 Human immune/haematopoietic antigen

47: LRTFQ KKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20}
48: RTFQK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
49: TFOKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
50: FOXXX KKKKKKKKKKKKKKKKKK KKKKK (K){20}
51: QKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKK XXG (K){20}

AAM90618 ck: 5691 len: 108 ! Aam90618 Human immune/haematopoietic antigen

78: VRPCL KKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20}
79: RPCLK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
80: PCLKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
81: CLKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
82: LKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}

[illegible]

$\{0, 1, 2, \dots, n\}$

1

87: KKKKK (K) {20} KKKKK
88: KKKKK (K) {20} KKKKK
89: KKKKK (K) {20} KKKKK
90: KKKKK (K) {20} KKKKK
91: KKKKK (K) {20} K
92: KKKKK (K) {20} KKKKK

AAM91933 ck: 8285 len: 55 ! Aam91933 Human digestive system antigen SEQ

1

33: PPTRP (R,K) {20,20} KKK
34: PTRPK (K) {20} KKK

AAM92433 ck: 1663 len: 87 ! Aam92433 Human digestive system antigen SEQ

1

36: KWSSX (R,K) {20,20} KKKKK
37: WSSXX (K) {20} KKKKK
38: SSYXX (K) {20} KKKKK
39: SXXXX (K) {20} KKKKK
40: XXXXX (K) {20} KKKKK
41: KKKKK (K) {20} KKKKK
42: KKKKK (K) {20} KKKKK
43: KKKKK (K) {20} KKKKK
44: KKKKK (K) {20} KKKKK
45: KKKKK (K) {20} KKKKK
46: KKKKK (K) {20} KKKKK
47: KKKKK (K) {20} KKKKK
48: KKKKK (K) {20} KKKKK
49: KKKKK (K) {20} KKKKK

50: KKKKK (K) {20} KKKKK
51: KKKKK (K) {20} KKKKK
52: KKKKK (K) {20} KKKKK
53: KKKKK (K) {20} KKKKK
54: KKKKK (K) {20} KKKKK
55: KKKKK (K) {20} KKKKK
56: KKKKK (K) {20} KKKKK
57: KKKKK (K) {20} KKKKK
58: KKKKK (K) {20} KKKKK
59: KKKKK (K) {20} KKKKK
60: KKKKK (K) {20} KKKKK
61: KKKKK (K) {20} KKKKK
62: KKKKK (K) {20} KKKKK
63: KKKKK (R,K) {20} KKKKK

AAO00092 ck: 9065 len: 113 ! Aao00092 Human polypeptide SEQ ID NO 13984.

1

91: CLGCL (R,K) {20,20} KFF
92: LGCLK (K) {20} PF

AAO00222 ck: 2916 len: 132 ! Aao00222 Human polypeptide SEQ ID NO 14114.

1

28: XPLPP (R,K) {20,20} GGPPK

AAO00232 ck: 1000 len: 102 ! Aao00232 Human polypeptide SEQ ID NO 14124.

1

29: RDCPF (R,K) {20,20} KKKKK
30: DCFEX (K) {20} KKKKK
31: CFFEX (K) {20} KKKKK

1
44: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKKKK NYFHV
AAO00608 ck: 7114 len: 57 ! Aao00608 Human polypeptide SEQ ID NO 14500.
22: SLSPK (R,K){20,20} KKKKKKKKKKKKKKKKKKKKK KSASS
23: LSPEK (K){20} KKKKKKKKKKKKKKKKKKKKK SASSS
AAO00619 ck: 6838 len: 117 ! Aao00619 Human polypeptide SEQ ID NO 14511.
1: KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
2: K KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
5: KKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
6: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
7: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
9: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
AAO01368 ck: 3955 len: 123 ! Aao01368 Human polypeptide SEQ ID NO 15260.
87: RQLIT (R,K){20,20} KKKKKKKKKKKKKKKKKKKKK KKKKK
88: QLITK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
89: LITKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
90: ITKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
91: TKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
92: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
93: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
94: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

95: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
96: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
97: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
98: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
AAO01560 ck: 7002 len: 74 ! Aao01560 Human polypeptide SEQ ID NO 15452
(R,K){20,20}
30: KLYHL KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: LYHLK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
32: YHLKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
33: HLKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
34: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
AAO02135 ck: 8265 len: 74 ! Aao02135 Human polypeptide SEQ ID NO 16027
(R,K){20,20}
41: MPPPP KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
42: PPPPK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
43: PPPPK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
44: PPPKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
45: PKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK

51: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKQ
52: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKQ
53: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KQ
54: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK Q

AAO02186 ck: 4844 len: 57 ! Aao02186 Human polypeptide SEQ ID NO 16078.

11: HCLLL (R,K){20,20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
12: CCLLK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK IKKK

AAO02310 ck: 704 len: 137 ! Aao02310 Human polypeptide SEQ ID NO 16202.

22: HSLNL (R,K){20,20} KKKKK KKKKKKKKKKKKKKKKKKK GGGVK

AAO02477 ck: 2018 len: 112 ! Aao02477 Human polypeptide SEQ ID NO 16369.

83: LASAV (R,K){20,20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
84: ASAVK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
85: SAVKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
86: AVKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
87: VKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
88: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
89: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKK
90: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKK
91: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KK
92: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK K
93: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK

AAO02733 ck: 2860 len: 132 ! Aao02733 Human polypeptide SEQ ID NO 16625.

86: FFFSL (R,K){20,20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

87: FFSLLK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
88: FSLKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
89: SLKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
90: LKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
91: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
92: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
93: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
94: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
95: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
96: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
97: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKA
98: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKAR
99: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKARS
100: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KARSR
101: KKKKK (R,K){20} KKKKK KKKKKKKKKKKKKKKKKKK ARSRS

AAO02946 ck: 3649 len: 126 ! Aao02946 Human polypeptide SEQ ID NO 16838.

25: DEATS (R,K){20,20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
26: EATSK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
27: ATSKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
28: TSKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
29: SKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
30: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
31: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
32: KKKKK (K){20} KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

33: KKKKK (K){20} KKKKK
34: KKKKK (K){20} KKKKK
35: KKKKK (K){20} KKKKK
36: KKKKK (K){20} KKKKK
37: KKKKK (K){20} KGGGP
38: KKKKK (K){20} GGGFK

AAO02961 ck: 4320 len: 83 ! Aao02961 Human polypeptide SEQ ID NO 16853.

22: RMPSS KKKKKKKKKKKKKKKKKKKKK KTAIT
23: MFSSK KKKKKKKKKKKKKKKKKKKKK TAITK

AAO03006 ck: 5212 len: 102 ! Aao03006 Human polypeptide SEQ ID NO 16998.

59: HDPFP KKKKKKKKKKKKKKKKKKKKK KKKKK
60: DFPFK KKKKKKKKKKKKKKKKKKKKK KKKKK
61: FFPKK KKKKKKKKKKKKKKKKKKKKK KKKKK
62: FPKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
63: PKKKK KKKKKKKKKKKKKKKKKKKKK KKKKA
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKAS
65: KKKKK KKKKKKKKKKKKKKKKKKKKK KTKASS
66: KKKKK KKKKKKKKKKKKKKKKKKKKK KASSS
67: KKKKK KKKKKKKKKKKKKKKKKKKKK ASSSS

AAO03024 ck: 6865 len: 64 ! Aao03024 Human polypeptide SEQ ID NO 16916.

15: SFAXA KKKKKKKKKKKKKKKKKKKKK KRGGP
16: FAXAR KKKKKKKKKKKKKKKKKKKKK RGGPP
17: AKARK KKKKKKKKKKKKKKKKKKKKK GGPLL

AAO03113 ck: 5587 len: 60 ! Aao03113 Human polypeptide SEQ ID NO 17005
33: EINML KKKKKKKKKKKKKKKKKKKKK KKKRG
34: INMLK KKKKKKKKKKKKKKKKKKKKK KRRGG
35: NMLKK KKKKKKKKKKKKKKKKKKKKK KRGGG
36: MLKKK KKKKKKKKKKKKKKKKKKKKK RGGGL
37: LKKKK KKKKKKKKKKKKKKKKKKKKK GGSL

AAO03123 ck: 6627 len: 108 ! Aao03123 Human polypeptide SEQ ID NO 17015

33: RTWRX KKKKKKKKKKKKKKKKKKKKK KKKKK
34: TWRXK KKKKKKKKKKKKKKKKKKKKK KKKKK
35: WRXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
36: RAXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
37: XKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKRG
40: KKKKK KKKKKKKKKKKKKKKKKKKKK KRRGG
41: KKKKK KKKKKKKKKKKKKKKKKKKKK KRRGG
42: KKKKK KKKKKKKKKKKKKKKKKKKKK RGGGQ
43: KKKKK (R,K){20} GGGQK

AAO03132 ck: 3903 len: 116 ! Aao03132 Human polypeptide SEQ ID NO 17024

1: KKKKKKKKKKKKKKKKKKKKK KKKKK
2: K KKKKKKKKKKKKKKKKKKKKK KKKRG
3: KK KKKKKKKKKKKKKKKKKKKKK KRRGG
4: KKK KKKKKKKKKKKKKKKKKKKKK KRRGP
5: KKKK KKKKKKKKKKKKKKKKKKKKK RGGPP

78: PSRAK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
79: SPAKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
80: RAKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
81: AKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
82: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
83: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
84: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKP
85: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKPG
86: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKPGG
87: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KPPGG
88: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK PGGGV
AAO03700 ck: 1279 len: 43 ! Aao03700 Human polypeptide SEQ ID NO 17592.
1 (R,K) {20,20}
(K) {20}
19: YSURL KKKKKKKKKKKKKKKKKKKKKKKKKKK GGGPF
AAO03703 ck: 7016 len: 113 ! Aao03703 Human polypeptide SEQ ID NO 17595.
1 (R,K) {20,20}
(K) {20}
58: IWDAA KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
59: WDAEK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
60: DAEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
61: AEKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKS
62: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKSP
63: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKSPG
64: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KSPGG
65: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK SPGGA

AAO03766 ck: 8808 len: 81 ! Aao03766 Human polypeptide SEQ ID NO 17658.
1 (R,K) {20,20}

49: TTTAC (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
50: TTACK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
51: TACKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
52: ACKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
53: CKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKG
54: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKGG
55: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKGGG
56: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KGGGG
57: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK GGGGA

AAO03841 ck: 8734 len: 100 ! Aao03841 Human polypeptide SEQ ID NO 17733.
1 (R,K) {20,20}
(K) {20}

36: KQHYP KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
37: QHYPK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
38: HYPKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
39: YPKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
40: PKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
41: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKRT
42: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKTK
43: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KRTKQ
44: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK RTKQK
45: KKKKK (R,K) {20} KKKKKKKKKKKKKKKKKKKKK TKQKK

AAO03906 ck: 4312 len: 100 ! Aao03906 Human polypeptide SEQ ID NO 17798.
1 (R,K) {20,20}
(K) {20}

34: NKQNG KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
35: KQNGK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
36: QNGKK KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK

59: KKKKK (K){20}
60: KKKKK (K){20}
61: KKKKK (K){20}
62: KKKKK (K){20}

AAO04647 ck: 4805 len: 58 ! Aao04647 Human polypeptide SEQ ID NO 18539.

28: KPTRP (R,K){20,20}
29: PTRPK (K){20}
30: TRPKK (K){20}
31: RPKKK (K){20}
32: PKKKK (K){20}
33: KKKKK (K){20}
34: KKKKK (K){20}
35: KKKKK (K){20}
36: KKKKK (K){20}
37: KKKKK (K){20}
38: KKKKK (K){20}

AAO04674 ck: 2036 len: 33 ! Aao04674 Human polypeptide SEQ ID NO 18566.

9: WCYIT (R,K){20,20}
10: CYITK (K){20}
11: YITKK (K){20}

AAO04679 ck: 7190 len: 31 ! Aao04679 Human polypeptide SEQ ID NO 18571.

8: CWFTQ (R,K){20,20}
9: WFTQK (K){20}

AAO04682 ck: 5605 len: 60 ! Aao04682 Human polypeptide SEQ ID NO 18574

2: L KKKKK (R,K){20,20}
3: LK KKKKK (K){20}
4: LKK KKKKK (K){20}
5: LKKK KKKKK (K){20}

AAO04690 ck: 7157 len: 81 ! Aao04690 Human polypeptide SEQ ID NO 18582

1: KKKKK (R,K){20,20}
2: K KKKKK (K){20}
3: KK KKKKK (K){20}
4: KKK KKKKK (K){20}
5: KKKK KKKKK (K){20}
6: KKKKK KKKKK (K){20}
7: KKKKK KKKKK (K){20}
8: KKKKK KKKKK (K){20}

AAO04715 ck: 6984 len: 35 ! Aao04715 Human polypeptide SEQ ID NO 18607.

8: LGSKD (R,K){20,20}
9: GSKDK KKKKK (K){20}
10: SKDKK KKKKK (K){20}

AAO04743 ck: 9412 len: 54 ! Aao04743 Human polypeptide SEQ ID NO 18635.

21: KINKL (R,K){20,20}
22: INKLK KKKKK (K){20}
23: NKLKK KKKKK (R,K){20}

AAO04747 ck: 8399 len: 39 ! Aao04747 Human polypeptide SEQ ID NO 18639.

1

(R,K){20,20}
(K){20}
13: RTGFV KKKKKKKKKKKKKKKKKKKKK KKG
(K){20}
14: TGFVK KKKKKKKKKKKKKKKKKKKKK KKG
(K){20}
15: GFVK KKKKKKKKKKKKKKKKKKKKK RGGF
(R,K){20}
16: FVKK KKKKKKKKKKKKKKKKKKKKK RGGF

AAO04752 ck: 1021 len: 71 ! Aao04752 Human polypeptide SEQ ID NO 18644.

1

(R,K){20,20}
(K){20}
19: QOGL KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
20: EOGLK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
21: OGLKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
22: GLKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
23: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKG
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKG
(K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKKK KGG
(K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKK GG

AAO04755 ck: 5521 len: 59 ! Aao04755 Human polypeptide SEQ ID NO 18647.

1

(R,K){20,20}
(K){20}
31: INSLE KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
32: NSLEK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
33: SLEKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
34: LEKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
35: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKT
(K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KT
(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKKKK T

AAO04756 ck: 6952 len: 26 ! Aao04756 Human polypeptide SEQ ID NO 18648

1

(R,K){20,20}
(K){20}
4: FFY KKKKKKKKKKKKKKKKKKKKK SSS

AAO04758 ck: 7836 len: 115 ! Aao04758 Human polypeptide SEQ ID NO 18650

1

(R,K){20,20}
(K){20}
7: PFYQL KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
8: FYQLK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
9: YQLKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
10: QLKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
11: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}

27: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
28: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
29: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
30: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
31: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
32: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
33: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
34: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
35: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
36: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
37: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
38: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
39: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
40: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
41: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
42: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
43: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
44: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
45: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
46: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
47: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
48: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
49: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK
50: KKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKKK

1

AAO04764 ck: 2035 len: 54 ! Aao04764 Human polypeptide SEQ ID NO 18656
(R,K) {20,20}
(K) {20}
9: GDSSL KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
10: DSSLK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
11: SSLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
12: SLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
13: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

AAO04802 ck: 1223 len: 105 ! Aao04802 Human polypeptide SEQ ID NO 18694
(R,K) {20,20}
(K) {20}
34: PFSRQ KKKKKKKKKKKKKKKKKKK GGGGL

1

AAO04854 ck: 4121 len: 58 ! Aao04854 Human polypeptide SEQ ID NO 18746.
(R,K) {20,20}
(K) {20}
13: TPFRA KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
14: PFRAK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
15: FRANK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
16: RAKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
17: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1

27: KKKKK (K) {20}
 28: KKKKK (K) {20}
 29: KKKKK (K) {20}
 30: KKKKK (K) {20}
 31: KKKKK (K) {20}
 32: KKKKK (R,K) {20}

AAO04856 ck: 4861 len: 135 ! Aao04856 Human polypeptide SEQ ID NO 18748.

41: LTLTT (R,K) {20,20}
 42: TLTTK (K) {20}
 43: LTTKK (K) {20}
 44: TTTKK (K) {20}
 45: TTKKK (K) {20}
 46: KKKKK (K) {20}
 47: KKKKK (K) {20}
 48: KKKKK (K) {20}
 49: KKKKK (K) {20}
 50: KKKKK (K) {20}
 51: KKKKK (K) {20}
 52: KKKKK (K) {20}
 53: KKKKK (K) {20}

AAO04872 ck: 3224 len: 93 ! Aao04872 Human polypeptide SEQ ID NO 18764.

8: FLYKL (R,K) {20,20}
 9: LYKLK (K) {20}

10: YKLKK (K) {20}
 11: KLKKK (R,K) {20}
 12: LKKKK (K) {20}

AAO04873 ck: 7719 len: 66 ! Aao04873 Human polypeptide SEQ ID NO 18765

21: SFLIE (R,K) {20,20}
 22: FLIEK (K) {20}

AAO04874 ck: 5753 len: 58 ! Aao04874 Human polypeptide SEQ ID NO 18766

10: SEKLP (R,K) {20,20}
 11: EKLPK (K) {20}
 12: KLPKK (K) {20}

AAO04881 ck: 8841 len: 115 ! Aao04881 Human polypeptide SEQ ID NO 18773

24: ITPHP (R,K) {20,20}
 25: TPHPK (K) {20}
 26: PPHPK (K) {20}
 27: HPPKK (K) {20}
 28: PPKKK (K) {20}
 29: KKKKK (K) {20}
 30: KKKKK (K) {20}
 31: KKKKK (K) {20}

AAO04917 ck: 3999 len: 103 ! Aao04917 Human polypeptide SEQ ID NO 18809

1: (R,K) {20,20}

AAO04928 ck: 23 len: 43 ! Aao04928 Human polypeptide SEQ ID NO 18820

19: PPKKE (R,K) {20,20}

AAO04969 ck: 4345 len: 57 ! Aao04969 Human polypeptide SEQ ID NO 18861.

1

(R,K){20,20}
(K){20}
26: LMGTG KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: MGTSK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: GTSKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: TSKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: SKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKKK K
(K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKKK

AAO05000 ck: 5847 len: 38 ! Aao05000 Human polypeptide SEQ ID NO 18892.

1

(R,K){20,20}
(K){20}
5: SGRS KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: SGRSK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: GRSKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: RSKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: SKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KRGG
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK RGGR
(R,K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK GGR

AAO05081 ck: 2260 len: 40 ! Aao05081 Human polypeptide SEQ ID NO 18973.

1

(R,K){20,20}
(K){20}
9: SRASP KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: RASPK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: ASPKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: SPKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: PKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKS
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKSS
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KSSSK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KSSK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK SSK

AAO05130 ck: 6597 len: 116 ! Aao05130 Human polypeptide SEQ ID NO 19022.

1

(R,K){20,20}
(K){20}
14: ETPPQ KKKKKKKKKKKKKKKKKKKKK KRGGG
(K){20}
15: TPFQK KKKKKKKKKKKKKKKKKKKKK RGGGF
(R,K){20}
16: PFOKK KKKKKKKKKKKKKKKKKKKKK GGGFL

AAO05191 ck: 4399 len: 85 ! Aao05191 Human polypeptide SEQ ID NO 19083.

1

(R,K){20,20}
(R,K){20}
30: GYIQL KKKKKKKKKKKKKKKKKKKKK GGAFK

AAO05260 ck: 3618 len: 55 ! Aao05260 Human polypeptide SEQ ID NO 19152.

1

(R,K){20,20}
(K){20}
15: PHKHQ KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: HKHKQ KKKKKKKKKKKKKKKKKKKKK KKKKK

[illegible][illegible]

AAO05372 ck: 9880 len: 43 ! Aao05372 Human polypeptide SEQ ID NO 19264.

1 AAO05665 ck: 432 len: 28 ! Aao05665 Human polypeptide SEQ ID NO 19557.
(R,K){20,20}
(K){20}
6: DFLQK KKKKKKKKKKKKKKKKKKKKKK RGG
(R,K){20}
7: FLQK KKKKKKKKKKKKKKKKKKKKKK GG
1 AAO06186 ck: 998 len: 88 ! Aao06186 Human polypeptide SEQ ID NO 20078.
(R,K){20,20}
(K){20}
28: SLLPX KKKKKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
29: LLPXK KKKKKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
30: LPXKK KKKKKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
31: PXKKK KKKKKKKKKKKKKKKKKKKKKK KGGGG
(K){20}
32: XKKKK KKKKKKKKKKKKKKKKKKKKKK GGGGF
1 AAO06357 ck: 4679 len: 52 ! Aao06357 Human polypeptide SEQ ID NO 20249.
(R,K){20,20}
(K){20}
9: FTCLI KKKKKKKKKKKKKKKKKKKKKK KKKIK
(K){20}
10: TCLIK KKKKKKKKKKKKKKKKKKKKKK KKKIK
(K){20}
11: CLIKK KKKKKKKKKKKKKKKKKKKKKK KTKKK
(K){20}
12: LIKKK KKKKKKKKKKKKKKKKKKKKKK IKKKK
1 AAO06429 ck: 2585 len: 71 ! Aao06429 Human polypeptide SEQ ID NO 20321.
(R,K){20,20}
(K){20}
29: VIIMX KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: IIMXK KKKKKKKKKKKKKKKKKKKKKK KKKKG
(K){20}
31: IMXKK KKKKKKKKKKKKKKKKKKKKKK KKKGR
(K){20}
32: MXKKK KKKKKKKKKKKKKKKKKKKKKK KKGKG
(K){20}
33: XKKKK KKKKKKKKKKKKKKKKKKKKKK KGRGG
(K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKKKK GRGGA
1 AAO06922 ck: 7296 len: 111 ! Aao06922 Human polypeptide SEQ ID NO 20814.
(R,K){20,20}
(K){20}
13: HSSEL KKKKKKKKKKKKKKKKKKKKKK GPPPK

1 AAO07241 ck: 3134 len: 40 ! Aao07241 Human polypeptide SEQ ID NO 21133.
(R,K){20,20}
(K){20}
4: YFP KKKKKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
5: YFPK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
6: YFPKK KKKKKKKKKKKKKKKKKKKKKK KKKKG
(R,K){20}
7: FPKKK KKKKKKKKKKKKKKKKKKKKKK KKKGG
(R,K){20}
8: PKKKK KKKKKKKKKKKKKKKKKKKKKK KKKGG
1 AAO07354 ck: 6513 len: 133 ! Aao07354 Human polypeptide SEQ ID NO 21346.
(R,K){20,20}
(K){20}
36: KXYHL KKKKKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
37: XYHLK KKKKKKKKKKKKKKKKKKKKKK KKKGP
(K){20}
38: YHLKK KKKKKKKKKKKKKKKKKKKKKK KGGPL
(K){20}
39: HLKKK KKKKKKKKKKKKKKKKKKKKKK GGPLK
1 AAO07410 ck: 693 len: 80 ! Aao07410 Human polypeptide SEQ ID NO 21302.
(R,K){20,20}
(K){20}
56: HKNPI KKKKKKKKKKKKKKKKKKKKKK RGGGF
(R,K){20}
57: KNPIK KKKKKKKKKKKKKKKKKKKKKK GGGF
1 AAO07412 ck: 4623 len: 166 ! Aao07412 Human polypeptide SEQ ID NO 21304.
(R,K){20,20}
(K){20}
93: VNTQX KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
94: NTQXK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
95: TQXKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
96: QXKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
97: XKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
98: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
99: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

101: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
102: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
103: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKGG
104: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKGG
105: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKGG
106: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKGG
107: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKGG

AAO07505 ck: 3528 len: 72 ! Aao07505 Human polypeptide SEQ ID NO 21397.

(R,K){20,20}

14: HLVEA KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
15: LVEAK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
16: VEAKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
17: EAKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
18: AKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
21: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
22: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
23: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

AAO07509 ck: 4998 len: 140 ! Aao07509 Human polypeptide SEQ ID NO 21401.

(R,K){20,20}

17: KVQXE KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
18: VKQXE KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
19: KQKXE KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
20: XEKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

21: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
25: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
26: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
27: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
28: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKGG
30: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKGG
31: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKGG
32: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKGG

AAO07594 ck: 7271 len: 93 ! Aao07594 Human polypeptide SEQ ID NO 21486.

(R,K){20,20}

9: FLILG KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
10: LLLOK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
11: LLGKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
12: LGKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
13: GKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

AAO07607 ck: 2434 len: 72 ! Aao07607 Human polypeptide SEQ ID NO 21499.

(R,K){20,20}

21: KKKKE KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
48: GLFYE KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

AAO07610 ck: 5270 len: 74 ! Aao07610 Human polypeptide SEQ ID NO 21502.

(R,K){20,20}

19: TELTI KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
20: ELTIK KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK

21: LTIKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
22: TIKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
23: IKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
25: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
26: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO07755 ck: 6069 len: 90 ! Aao07755 Human polypeptide SEQ ID NO 21647.

35: KNQSS KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
36: NQSSK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
37: QSSKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
38: SSKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
39: SKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
40: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO07762 ck: 5115 len: 41 ! Aao07762 Human polypeptide SEQ ID NO 21654.

18: IPSLX KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
19: PSLXK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
20: SLXKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
21: LXXKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
22: XKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO07775 ck: 9827 len: 90 ! Aao07775 Human polypeptide SEQ ID NO 21667.

46: PHXPE KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}

47: HXPEK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
48: XPEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
49: PEKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
50: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO07806 ck: 1959 len: 99 ! Aao07806 Human polypeptide SEQ ID NO 21698

41: KLHCI KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
42: LHCIK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
43: HCIKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
44: CIKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
45: IKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
46: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
47: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
48: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO07872 ck: 8003 len: 135 ! Aao07872 Human polypeptide SEQ ID NO 21764

52: TKKKQ KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}

AAO07874 ck: 8007 len: 111 ! Aao07874 Human polypeptide SEQ ID NO 21766

90: BEPSI KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
91: EPSIK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO07876 ck: 8130 len: 66 ! Aao07876 Human polypeptide SEQ ID NO 21768

17: IYSFE KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
18: YSFEK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
19: SPEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

20: FEKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKQK
21: EKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKQK
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKE
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK QKKEN

AAO07976 ck: 6471 len: 97 ! Aao07976 Human polypeptide SEQ ID NO 21868.

43: FPIST (R,K){20,20} (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
44: PISTK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
45: ISTKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
46: STKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
47: TKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
48: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
49: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
50: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
51: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKA
52: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKAS
53: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKAAS
54: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KASSG
55: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK ASSGG

AAO07981 ck: 9551 len: 81 ! Aao07981 Human polypeptide SEQ ID NO 21873.

54: KPCIY (R,K){20,20} (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
55: PCIYK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
56: CIYKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
57: IYKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
58: YKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

59: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKK
60: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KK
61: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK K
62: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK

AAO08018 ck: 9768 len: 48 ! Aao08018 Human polypeptide SEQ ID NO 21910.

24: KISQV (R,K){20,20} (R,K){20} KKKKK
25: ISQVR (K){20} KKKKKKKKKKKKKKKKKKK KKKK
26: SQVRK (K){20} KKKKKKKKKKKKKKKKKKK KKK
27: QVRKK (K){20} KKKKKKKKKKKKKKKKKKK KK
28: VRKKK (K){20} KKKKKKKKKKKKKKKKKKK K
29: RKKKK (K){20} KKKKKKKKKKKKKKKKKKK

AAO08121 ck: 5381 len: 86 ! Aao08121 Human polypeptide SEQ ID NO 22013.

20: HFGLL (R,K){20,20} (K){20} KKKKKKKKKKKKKKKKKKK GGGAF

AAO08232 ck: 782 len: 104 ! Aao08232 Human polypeptide SEQ ID NO 22124.

49: ISQQN (R,K){20,20} (K){20} KKKKKKKKKKKKKKKKKKK KDGSG
50: SQQNK (K){20} KKKKKKKKKKKKKKKKKKK DGGGA

AAO08450 ck: 6228 len: 31 ! Aao08450 Human polypeptide SEQ ID NO 22342.

5: NNSA (R,K){20,20} (K){20} KKKKKKKKKKKKKKKKKKK KKGGA
6: NNSAK (K){20} KKKKKKKKKKKKKKKKKKK KGGAA
7: NSAKK (K){20} KKKKKKKKKKKKKKKKKKK GGAAP

AAO08458 ck: 1813 len: 59 ! Aao08458 Human polypeptide SEQ ID NO 22350.

8: LGOME (R,K){20,20} (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

9: GWEK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
10: GWEK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
11: WEKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
12: EKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
13: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
AAO08460 ck: 5334 len: 62 ! Aao08460 Human polypeptide SEQ ID NO 22352.
1
(R,K) {20,20}
(K) {20}
22: KEFST KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
23: EFSTK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
24: FSTKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
25: STKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
26: TKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
32: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
33: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
34: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
AAO08465 ck: 8728 len: 107 ! Aao08465 Human polypeptide SEQ ID NO 22358.
1
(R,K) {20,20}
(K) {20}
57: FWGGV KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
58: WGGVK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
59: GGVKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}
60: GVKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K) {20}

61: VTKKK KKKKKKKKKKKKKKKKKKKKK KRRKG
(K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKK RKGR
(R,K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK RKGRK
(R,K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KGRKE
(R,K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK GRKEG
(R,K) {20,20}
(K) {20}
AAO08580 ck: 5164 len: 62 ! Aao08580 Human polypeptide SEQ ID NO 22472
1
(R,K) {20,20}
(K) {20}
30: PILLL KKKKKKKKKKKKKKKKKKKKK GGALK
(R,K) {20,20}
(K) {20}
AAO08591 ck: 900 len: 46 ! Aao08591 Human polypeptide SEQ ID NO 22483
1
(R,K) {20,20}
(K) {20}
14: LLCSS KKKKKKKKKKKKKKKKKKKKK GGAFK
(K) {20}
AAO08609 ck: 5151 len: 74 ! Aao08609 Human polypeptide SEQ ID NO 22501
1
(R,K) {20,20}
(K) {20}
18: KCVIL KKKKKKKKKKKKKKKKKKKKK KKKKL
(K) {20}
19: CVILK KKKKKKKKKKKKKKKKKKKKK GKLLK
(R,K) {20,20}
(K) {20}
AAO08612 ck: 557 len: 46 ! Aao08612 Human polypeptide SEQ ID NO 22504
1
(R,K) {20,20}
(K) {20}
20: TFCIM KKKKKKKKKKKKKKKKKKKKK KKKGG
(K) {20}
21: FCIMK KKKKKKKKKKKKKKKKKKKKK KKGGA
(K) {20}
22: CIMKK KKKKKKKKKKKKKKKKKKKKK KGGAL
(K) {20}
23: IMKKK KKKKKKKKKKKKKKKKKKKKK GGAL
(R,K) {20,20}
(K) {20}
AAO08616 ck: 4555 len: 41 ! Aao08616 Human polypeptide SEQ ID NO 22508
1
(R,K) {20,20}
(K) {20}
6: HCALP KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
7: CALPK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
8: ALPKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
9: LPKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
10: PKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

11: KKKKK (K) {20} KKKKK
 12: KKKKK (K) {20} KKKKK
 13: KKKKK (K) {20} KKKKK
 14: KKKKK (K) {20} KKKKK
 15: KKKKK (K) {20} KKKKK
 16: KKKKK (K) {20} KKKKK
 17: KKKKK (K) {20} KKKKK
 18: KKKKK (K) {20} KKKKK
 19: KKKKK (K) {20} KKKKK

AAO08653 ck: 6448 len: 63 ! Aao08653 Human polypeptide SEQ ID NO 22545.

12: PIKLT (R,K) {20,20} KKKKK
 13: IKLTK (K) {20} KKKKK
 14: KLTKK (K) {20} KKKKK
 15: LTKKK (K) {20} KKKKK
 16: TKKKK (K) {20} KKKKK
 17: KKKKK (K) {20} KKKKK
 18: KKKKK (K) {20} KKKKK
 19: KKKKK (R,K) {20} KKKKK
 20: KKKKK (K) {20} KKKKK

AAO08686 ck: 6936 len: 81 ! Aao08686 Human polypeptide SEQ ID NO 22578.

23: HPVNH (R,K) {20,20} KKKKK
 24: PVNHK (K) {20} KKKKK
 25: VNHKK (K) {20} KKKKK
 26: NHKKK (K) {20} KKKKK

27: HKKKK (K) {20} KKKKK
 28: KKKKK (K) {20} KKKKK
 29: KKKKK (K) {20} KKKKK
 30: KKKKK (K) {20} KKKKK
 31: KKKKK (K) {20} KKKKK
 32: KKKKK (K) {20} KKKKK
 33: KKKKK (K) {20} KKKKK
 34: KKKKK (K) {20} KKKKK
 35: KKKKK (K) {20} KKKKK
 36: KKKKK (K) {20} KKKKK
 37: KKKKK (K) {20} KKKKK

AAO08701 ck: 588 len: 46 ! Aao08701 Human polypeptide SEQ ID NO 22593.

7: INANS (R,K) {20,20} KKKKK
 8: NANSK (K) {20} KKKKK
 9: ANSKK (K) {20} KKKKK
 10: NSKKK (K) {20} KKKKK
 11: SKKKK (K) {20} KKKKK
 12: KKKKK (K) {20} KKKKK
 13: KKKKK (K) {20} KKKKK

AAO08707 ck: 9359 len: 48 ! Aao08707 Human polypeptide SEQ ID NO 22599.

13: TAVTQ (R,K) {20,20} KKKKK
 14: AVTQK (K) {20} KKKKK
 15: VTQKK (K) {20} KKKKK

1 AAO08708 ck: 9654 len: 32 ! Aao08708 Human polypeptide SEQ ID NO 22600.
(R,K){20,20}
(K){20}
8: SMTFP KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: MTFPK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: TFPKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
11: PPKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
12: PPKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAO08717 ck: 4688 len: 99 ! Aao08717 Human polypeptide SEQ ID NO 22609.
(R,K){20,20}
(K){20}
80: SFLLI KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
1 AAO08772 ck: 1431 len: 105 ! Aao08772 Human polypeptide SEQ ID NO 22664.
(R,K){20,20}
(K){20}
23: IMNYL KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
24: MNYLK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
25: NYLKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
26: YLEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
27: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAO08774 ck: 2529 len: 71 ! Aao08774 Human polypeptide SEQ ID NO 22666.
(R,K){20,20}
(K){20}
16: RTIKL KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
17: TIKLK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
18: IKLKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
19: KLEKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

20: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAO08817 ck: 529 len: 28 ! Aao08817 Human polypeptide SEQ ID NO 22709.
(R,K){20,20}
(K){20}
3: AX KKKKKKKKKKKKKKKKKKKKKKKKKKK NIW
(K){20}
AAO08820 ck: 1888 len: 84 ! Aao08820 Human polypeptide SEQ ID NO 22712
(R,K){20,20}
(K){20}
28: CMTFS KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
29: MTFSK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
AAO08841 ck: 3529 len: 74 ! Aao08841 Human polypeptide SEQ ID NO 22733.
(R,K){20,20}
(K){20}
3: PQ KKKKKKKKKKKKKKKKKKKKKKKKKKK KWGG
(K){20}
4: PQK KKKKKKKKKKKKKKKKKKKKKKKKKKK WGGF
(K){20}
AAO08857 ck: 2816 len: 49 ! Aao08857 Human polypeptide SEQ ID NO 22749.
(R,K){20,20}
(K){20}
25: PPPTS KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
26: PPTSK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
27: PTSKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
28: TSKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
29: SKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAO08913 ck: 3132 len: 52 ! Aao08913 Human polypeptide SEQ ID NO 22805.
(R,K){20,20}
(K){20}
9: IIKSF KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: IKSFK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
11: KSFKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

4: AQQ KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
5: AQQK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
6: AQQKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
7: QQQKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
8: QKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO09072 ck: 8432 len: 42 ! Aao09072 Human polypeptide SEQ ID NO 22964.

11: LIPTL KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
12: IFTLK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
13: FTLKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
14: TLKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO09077 ck: 9524 len: 87 ! Aao09077 Human polypeptide SEQ ID NO 22969.

26: RPLLT KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
27: FLLTK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
28: LLTKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
29: LTKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
30: TKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO09162 ck: 3841 len: 100 ! Aao09162 Human polypeptide SEQ ID NO 23054.

11: KTLFQ KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
12: TLFQK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
13: LFQKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
14: FQKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

15: QKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
23: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
25: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
26: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO09258 ck: 7988 len: 42 ! Aao09258 Human polypeptide SEQ ID NO 23150.

7: LLFCP KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}
8: LFCPK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
9: FCPKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
10: CPKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
11: PKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK (K) {20}

AAO09269 ck: 7553 len: 106 ! Aao09269 Human polypeptide SEQ ID NO 23161.

39: NTGML KKKKKKKKKKKKKKKKKKKKKKKKKKK (R,K) {20,20}

AAO09457 ck: 9658 len: 32 ! Aao09457 Human polypeptide SEQ ID NO 23349.
(R,K){20,20}
(K){20}
9: ALVPQ KKKKKKKKKKKKKKKKKKKKK NIKI

AAO09819 ck: 4709 len: 80 ! Aao09819 Human polypeptide SEQ ID NO 23711.
(R,K){20,20}
(K){20}
54: CFFVX KKKKKKKKKKKKKKKKKKKKK EKGGG

AAO10447 ck: 6342 len: 50 ! Aao10447 Human polypeptide SEQ ID NO 24339.
(R,K){20,20}
(K){20}
25: RPLKL KKKKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
26: FLKLK KKKKKKKKKKKKKKKKKKKKK KKKKI
(R,K){20}
27: LKLKK KKKKKKKKKKKKKKKKKKKKK KKKI
(R,K){20}
28: KLKKK KKKKKKKKKKKKKKKKKKKKK KKI
(R,K){20}
29: LKKKK KKKKKKKKKKKKKKKKKKKKK KI
(R,K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK I

AAO10451 ck: 3955 len: 82 ! Aao10451 Human polypeptide SEQ ID NO 24343.
(R,K){20,20}
(K){20}
59: SRASP KKKKKKKKKKKKKKKKKKKKK ARGG

AAO10467 ck: 7542 len: 116 ! Aao10467 Human polypeptide SEQ ID NO 24359.
(R,K){20,20}
(K){20}
57: KCEFM KKKKKKKKKKKKKKKKKKKKK KIGGG
(K){20}
58: CEFMK KKKKKKKKKKKKKKKKKKKKK IGGA

AAO10564 ck: 9156 len: 98 ! Aao10564 Human polypeptide SEQ ID NO 24456.
(R,K){20,20}
(K){20}
18: NLLTL KKKKKKKKKKKKKKKKKKKKK KCEMP
(K){20}
19: LLTLK KKKKKKKKKKKKKKKKKKKKK KEMPV
(K){20}
20: LTLKK KKKKKKKKKKKKKKKKKKKKK EMPVK

AAO10608 ck: 9106 len: 67 ! Aao10608 Human polypeptide SEQ ID NO 24500.
(R,K){20,20}
(K){20}
42: CRLSE KKKKKKKKKKKKKKKKKKKKK KKKKD

(K){20}
43: RLSEK KKKKKKKKKKKKKKKKKKKKK KKDD
(K){20}
44: LSEKK KKKKKKKKKKKKKKKKKKKKK KKDS
(K){20}
45: SEKKK KKKKKKKKKKKKKKKKKKKKK KDS
(K){20}
46: EKKKK KKKKKKKKKKKKKKKKKKKKK DS

AAO10638 ck: 621 len: 56 ! Aao10638 Human polypeptide SEQ ID NO 24530.
(R,K){20,20}
(K){20}
33: CEPQP KKKKKKKKKKKKKKKKKKKKK KARG
(K){20}
34: EPQPK KKKKKKKKKKKKKKKKKKKKK ARG

AAO10786 ck: 7349 len: 65 ! Aao10786 Human polypeptide SEQ ID NO 24678.
(R,K){20,20}
(K){20}
24: CPXFS KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
25: PXFSK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
26: XFSKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
27: FSKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
28: SKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKV
(K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKVF
(K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKKK KKVFF
(K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKKK KVFFF
(K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKKK VFFE

AAO10852 ck: 5147 len: 41 ! Aao10852 Human polypeptide SEQ ID NO 24744.
(R,K){20,20}
(K){20}
20: PSRXM KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
21: SRXMK KKKKKKKKKKKKKKKKKKKKK K
(K){20}

22: RMMKK KKKKKKKKKKKKKKKKKKK

AAO10853 ck: 444 len: 98 ! Aao10853 Human polypeptide SEQ ID NO 24745.

1

(R,K){20,20}

14: LRAHL KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

15: RAHLK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

16: AHLKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

17: HLAKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

18: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AAO10859 ck: 2690 len: 70 ! Aao10859 Human polypeptide SEQ ID NO 24751.

1

(R,K){20,20}

44: XGITE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: GITEK KKKKKKKKKKKKKKKKKKK KKKKK

AAO10933 ck: 849 len: 69 ! Aao10933 Human polypeptide SEQ ID NO 24825.

1

(R,K){20,20}

33: FINTE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

34: INTEK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

35: NTEKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

36: TEKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

37: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AAO10973 ck: 824 len: 51 ! Aao10973 Human polypeptide SEQ ID NO 24865.

1

(R,K){20,20}

(R,K){20}

16: SYFXM KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

17: YFXMR KKKKKKKKKKKKKKKKKKK KKKKK

AAO10983 ck: 694 len: 51 ! Aao10983 Human polypeptide SEQ ID NO 24875.

1

(R,K){20,20}

(K){20}

24: HPILX KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: PILXX KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

26: ILXXK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

27: LXXKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

28: XXXKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKK TGG

AAO10995 ck: 2382 len: 122 ! Aao10995 Human polypeptide SEQ ID NO 24887.

1

(R,K){20,20}

(K){20}

22: PCXGS KKKKKKKKKKKKKKKKKKK KKKKK

23: CKGSK KKKKKKKKKKKKKKKKKKKKK KKKK (K){20}
 24: KGSKK KKKKKKKKKKKKKKKKKKKKK KKKKG (K){20}
 25: GSKKK KKKKKKKKKKKKKKKKKKKKK KKKGG (K){20}
 26: SKKKK KKKKKKKKKKKKKKKKKKKKK KGGGG (K){20}
 27: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGGG (K){20}
 28: KKKKK KKKKKKKKKKKKKKKKKKKKK GGGGF (K){20}

AAO10997 ck: 5315 len: 58 ! Aao10997 Human polypeptide SEQ ID NO 24889.

29: IXIIQ KKKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20}
 30: XIIOK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 31: IIQKK KKKKKKKKKKKKKKKKKKKKK KKKRG (K){20}
 32: IQKKK KKKKKKKKKKKKKKKKKKKKK KRGGG (K){20}
 33: QKKKK KKKKKKKKKKKKKKKKKKKKK KRGGP (K){20}
 34: KKKKK KKKKKKKKKKKKKKKKKKKKK RGPPP (R,K){20}
 35: KKKKK KKKKKKKKKKKKKKKKKKKKK GGGP (K){20}

AAO11002 ck: 8638 len: 82 ! Aao11002 Human polypeptide SEQ ID NO 24894.

5: FXKRS KKKKKKKKKKKKKKKKKKKKK GGALL (R,K){20,20}
 38: XFKRS KKKKKKKKKKKKKKKKKKKKK GGALL (R,K){20}

AAO11033 ck: 3780 len: 49 ! Aao11033 Human polypeptide SEQ ID NO 24925.

10: QXKFI KKKKKKKKKKKKKKKKKKKKK KLSKK (R,K){20,20}
 11: XKFIR KKKKKKKKKKKKKKKKKKKKK LSKKI (K){20}

AAO11048 ck: 7868 len: 85 ! Aao11048 Human polypeptide SEQ ID NO 24940.

53: IISDP KKKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20}
 54: ISDPK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}

55: SDPKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 56: DPKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 57: PKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 64: KKKKK KKKKKKKKKKKKKKKKKKKKK KK (K){20}
 65: KKKKK KKKKKKKKKKKKKKKKKKKKK K (K){20}
 66: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20}

AAO11078 ck: 4186 len: 100 ! Aao11078 Human polypeptide SEQ ID NO 24970

39: PPKPD KKKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20,20}
 40: PKPDK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 41: KPDKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 42: PDKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 43: DKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 45: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 46: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 47: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
 48: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKRG (K){20}
 49: KKKKK KKKKKKKKKKKKKKKKKKKKK KRRGG (K){20}

50: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KRGGG
 51: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK RGGGG
 52: KKKKK (R,K) {20} KKKKKKKKKKKKKKKKKKKKK GGGGF
 AAO11124 ck: 5121 len: 61 ! Aao11124 Human polypeptide SEQ ID NO 25016.
 (R,K) {20,20}
 16: KYSFL KKKKKKKKKKKKKKKKKKKKK KKKKK
 17: YSFLK KKKKKKKKKKKKKKKKKKKKK KKKKK
 18: SFLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 19: FLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 20: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 26: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 27: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 31: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 32: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 33: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 34: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 35: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

1

36: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 37: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 38: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 39: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKK
 40: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KK
 41: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK K
 42: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 AAO11139 ck: 8807 len: 68 ! Aao11139 Human polypeptide SEQ ID NO 25031.
 (R,K) {20,20}
 22: VCRFP KKKKKKKKKKKKKKKKKKKKK KMBGG
 23: CRFPK KKKKKKKKKKKKKKKKKKKKK MBGGG
 AAO11165 ck: 7203 len: 62 ! Aao11165 Human polypeptide SEQ ID NO 25057.
 (R,K) {20,20}
 18: DASMV KKKKKKKKKKKKKKKKKKKKK KKKKK
 19: ASMVK KKKKKKKKKKKKKKKKKKKKK KKKKK
 20: SMVKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 21: MVKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 22: VKKKK KKKKKKKKKKKKKKKKKKKKK KKKKA
 23: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKKAS
 24: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KKAASS
 25: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK KASSS
 26: KKKKK (K) {20} KKKKKKKKKKKKKKKKKKK ASSSQ
 AAO11209 ck: 6562 len: 96 ! Aao11209 Human polypeptide SEQ ID NO 25101.
 (R,K) {20,20}
 63: LLLAY KKKKKKKKKKKKKKKKKKKKK KKKPKY
 64: LLAYK KKKKKKKKKKKKKKKKKKKKK KPKYL
 (K) {20}

1

1

1


```

1
5: KKKK KKKKKKKKKKKKKKKKKKKKK AGGGA
      (K) {20}
AAO11843 ck: 86 len: 36 ! Aao11843 Human polypeptide SEQ ID NO 25735.
1: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20,20}
      (K) {20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKSG
      (K) {20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKSQ
      (K) {20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KSG
      (K) {20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK SG
      (K) {20}
AAO11844 ck: 3574 len: 49 ! Aao11844 Human polypeptide SEQ ID NO 25736.
18: ILYNE KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20,20}
      (K) {20}
19: LYNEK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
20: YMEKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
21: MEKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
22: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}

```

```

23: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKYS
      (K) {20}
25: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKYSR
      (K) {20}
26: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KYSR
      (K) {20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK YSR
      (K) {20}
AAO11845 ck: 8375 len: 45 ! Aao11845 Human polypeptide SEQ ID NO 25737.
1
15: SPSTL KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20,20}
      (K) {20}
16: PSTLK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
17: STLKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
18: TLKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
19: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (K) {20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKT
      (K) {20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKTG
      (K) {20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKTG
      (K) {20}
23: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KTG
      (K) {20}
24: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK TG
      (K) {20}
AAO11849 ck: 284 len: 36 ! Aao11849 Human polypeptide SEQ ID NO 25741.
1
13: KCLCE KKKKKKKKKKKKKKKKKKKKKKKKK KKKR
      (R,K) {20,20}
      (K) {20}
14: CLCEK KKKKKKKKKKKKKKKKKKKKKKKKK RKR
      (K) {20}
15: LCEKK KKKKKKKKKKKKKKKKKKKKKKKKK KR
      (R,K) {20}
16: CEKKK KKKKKKKKKKKKKKKKKKKKKKKKK R
      (R,K) {20}
17: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
      (R,K) {20}
AAO11886 ck: 5282 len: 97 ! Aao11886 Human polypeptide SEQ ID NO 25778.
1
32: SNSSL KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
      (R,K) {20,20}
      (K) {20}

```

33: NSSLK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 34: SSLKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 35: SLKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 36: LKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 37: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 38: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 39: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 40: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 41: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 42: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 43: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 44: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 45: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 46: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK

AAO11888 ck: 6488 len: 31 ! Aao11888 Human polypeptide SEQ ID NO 25780.

2: D KKKKKKKKKKKKKKKKKKKKK KAGGG
 3: DK KKKKKKKKKKKKKKKKKKKKK AGGGL

AAO11903 ck: 8876 len: 48 ! Aao11903 Human polypeptide SEQ ID NO 25795.

17: VIKYV (R,K){20,20} (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 18: IKYVK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 19: KYVKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 20: YVKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 21: VKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK

22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 25: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 26: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 27: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 28: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK

AAO11999 ck: 3954 len: 47 ! Aao11999 Human polypeptide SEQ ID NO 25891.

17: IDBGL (R,K){20,20} (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 18: DEGLK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 19: EGLKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 20: GLKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 21: LKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK

AAO12098 ck: 8434 len: 39 ! Aao12098 Human polypeptide SEQ ID NO 25990.

13: TCRFG (R,K){20,20} (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 14: CKFGK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 15: KFGKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 16: FGKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 17: GKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
 18: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK

39: KKKKK (K){20} KKKSG
40: KKKKK (K){20} KKS GG
41: KKKKK (K){20} KSGGG
42: KKKKK (K){20} SGGGA

AAO12447 ck: 5017 len: 47 ! Aao12447 Human polypeptide SEQ ID NO 26339.

14: HLVCE (R,K){20,20} KRGGA
15: LVCEK (K){20} RGGAL
16: VCEKK (R,K){20} GGALK

AAO12476 ck: 837 len: 104 ! Aao12476 Human polypeptide SEQ ID NO 26368.

65: RFCHQ (R,K){20,20} KGFFF
66: FCHQK (K){20} GFFFW

AAO12548 ck: 6973 len: 74 ! Aao12548 Human polypeptide SEQ ID NO 26440.

38: AVLPL (R,K){20,20} KKG GG
39: VLPLK (K){20} KGGGV
40: LPLKK (K){20} GGVVF

AAO12553 ck: 1903 len: 33 ! Aao12553 Human polypeptide SEQ ID NO 26445.

9: SCCFI (R,K){20,20} KRGAP
10: CCFIK (K){20} RGAP
11: CFIKK (R,K){20} GAP

AAO13164 ck: 2798 len: 71 ! Aao13164 Human polypeptide SEQ ID NO 27056.

35: RPPLX (R,K){20,20} KEMFK
36: PPLXR (K){20} EMFKR

AAO13576 ck: 4846 len: 99 ! Aao13576 Human polypeptide SEQ ID NO 27468.

1

53: TNNLI (R,K){20,20} KKKKK
54: NNLIK (K){20} KKKKK
55: NLIKK (K){20} KKKKK
56: LIKKK (K){20} KKKKK
57: IKKKK (K){20} KKKGG
58: KKKKK (K){20} KKG GG
59: KKKKK (K){20} KGGGP
60: KKKKK (K){20} GGPPF

AAO13785 ck: 6241 len: 100 ! Aao13785 Human polypeptide SEQ ID NO 27677.

1

41: LYAPP (R,K){20,20} KKKKK
42: YAPPK (K){20} KKKKK
43: APPKK (K){20} KKKKK
44: PPKKK (K){20} KKKKK
45: PKKKK (K){20} KKKKK
46: KKKKK (K){20} KKKKK
47: KKKKK (K){20} KKKKK
48: KKKKK (K){20} KKKKK
49: KKKKK (K){20} KKKKK
50: KKKKK (K){20} KKKKK
51: KKKKK (K){20} KKKKK
52: KKKKK (K){20} KKKKK
53: KKKKK (K){20} KKKKA
54: KKKKK (K){20} KKKAG

55: KKKKK KKKKKKKKKKKKKKKKK KAGG

(K){20}

56: KKKKK KKKKKKKKKKKKKKKKK KAGG

(K){20}

57: KKKKK KKKKKKKKKKKKKKKKK AGGG

AAU17983 ck: 7016 len: 315 ! Aau17983 Human immunoglobulin polypeptide S

(R,K){20,20}

(R,K){20}

273: QVFAP RKKKKKKKKKKKKKKKKK KGGRS

(K){20}

274: VFAPR KKKKKKKKKKKKKKKKK KGGRS

AAU18049 ck: 7611 len: 315 ! Aau18049 Human immunoglobulin polypeptide S

(R,K){20,20}

(R,K){20}

273: QVFAP RKKKKKKKKKKKKKKKKK KGGRS

(K){20}

274: VFAPR KKKKKKKKKKKKKKKKK KGGRS

AAM14119 ck: 5383 len: 86 ! Aam14119 Peptide #553 encoded by probe for

(R,K){20,20}

(R,K){20}

15: RRRRG RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

16: RRRGR RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

17: RRRGR RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

18: RRRRR RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

19: GRRRR RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

20: RRRRR RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

21: RRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

22: RRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

23: RRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

24: RRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

25: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

26: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

27: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

28: KRRRK RRRKKKKKKKKKKKKKK KKRRR

29: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

30: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

31: RRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

32: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

33: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

34: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

35: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

36: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R){20}

37: RRRRK RRRKKKKKKKKKKKKKK KKRRR

(R){20}

38: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R){20}

39: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R){20}

40: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R){20}

41: KRRRK RRRKKKKKKKKKKKKKK KKRRR

(R){20}

42: RRRRK RRRKKKKKKKKKKKKKK KKRRR

(R){20}

43: RRRRK RRRKKKKKKKKKKKKKK KKRRR

(R){20}

44: RRRRK RRRKKKKKKKKKKKKKK KKRRR

(R){20}

45: RRRRK RRRKKKKKKKKKKKKKK KKRRR

AAM14961 ck: 1334 len: 86 ! Aam14961 Peptide #1395 encoded by probe for

(R,K){20,20}

(R,K){20}

57: EEEEG RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

58: EEEGR RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

59: EEEGR RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

60: EEEGR RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

61: GRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

62: RRRRK RRRKKKKKKKKKKKKKK KKRRR

(R,K){20}

63: RRRKK (R,K){20} KRRKKKKKKKKKKKKKKKKKK KKKK
64: RKKKK (R,K){20} RRRKKKKKKKKKKKKKKKKKK KKK
65: KKKKR (R,K){20} KRRKKKKKKKKKKKKKKKK KK
66: KKKRK (R,K){20} RRRKKKKKKKKKKKKKKKK K
67: KKKR (K){20} KKKKKKKKKKKKKKKKKKK

AAM15038 ck: 9082 len: 167 ! Aam15038 Peptide #1472 encoded by probe for

33: EGRG (R,K){20,20} RRRRRKKRRRRKKRRK RGGG
34: EGRG (R,K){20} RRRRRKKRRRRKKRRK RGGG
35: GRGR (R,K){20} RRRRRKKRRRRKKRRK RGGG

AAM15527 ck: 2276 len: 89 ! Aam15527 Peptide #1961 encoded by probe for

23: EEEE (R,K){20,20} KKKKKKKKKKKKKKKKK EEEK
46: KKEE (K){20} KKKKKKKKKKKKKKK KKEE
47: KEEK (K){20} KKKKKKKKKKKKKKK KKEE
48: BEEK (R,K){20} KKKKKKKKKKKKKKK KEEB
49: BEKK (R,K){20} KKKKKKKKKKKKKKK EEEE

AAM15826 ck: 1939 len: 130 ! Aam15826 Peptide #2260 encoded by probe for

42: EGRK (R,K){20,20} RRRRRRRRRRRRRKK RRRR
43: GRKR (R,K){20} RRRRRRRRRRRRRKK RRRR
44: RKRR (R,K){20} RRRRRRRRRRRRRKK RRRR
45: KERR (R,K){20} RRRRRRRRRRRRRKK RRRK
46: ERRR (R,K){20} RRRRRRRRRRRRRKK RRRK
47: RRRR (R,K){20} RRRRRRRRRRRRRKK RRRK
48: RRRR (R,K){20} RRRRRRRRRRRRRKK RRRK

49: RRRR (R,K){20} RRRRRRRRRRRRRKK KKKK
50: RRRR (R,K){20} RRRRRRRRRRRRRKK KKKK
51: RRRR (R,K){20} RRRRRRRRRRRRRKK KKEE
52: RRRR (R,K){20} RRRRRRRRRRRRRKK KKEE
53: RRRR (R,K){20} RRRRRRRRRRRRRKK KEEB
54: RRRR (R,K){20} RRRRRRRRRRRRRKK EEEE

AAM16123 ck: 3607 len: 88 ! Aam16123 Peptide #2557 encoded by probe for

39: ERKE (R,K){20,20} KKKKKKKKKKKKKKK KKKK
40: RKRE (K){20} KKKKKKKKKKKKKKK KKKK
41: KREN (K){20} KKKKKKKKKKKKKKK KKKK
42: REKK (K){20} KKKKKKKKKKKKKKK KKKK
43: EKKK (K){20} KKKKKKKKKKKKKKK KKKK
44: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
45: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
46: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
47: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
48: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
49: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
50: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
51: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
52: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
53: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
54: KKKK (K){20} KKKKKKKKKKKKKKK KKKK
55: KKKK (K){20} KKKKKKKKKKKKKKK KKKK

56: KKKKK (K){20} KKKKK
57: KKKKK (K){20} KKKKK
58: KKKKK (K){20} KKKKK
59: KKKKK (K){20} KKKKK
60: KKKKK (K){20} KKKKK
61: KKKKK (K){20} KKKKK
62: KKKKK (K){20} KKKKK
63: KKKKK (K){20} KKKKK
64: KKKKK (K){20} KKKKK
65: KKKKK (K){20} KKKKK
66: KKKKK (K){20} KKKKK
67: KKKKK (K){20} KKKKK
68: KKKKK (K){20} K
69: KKKKK (K){20} KKKKK

AAM16691 ck: 3937 len: 85 | Aam16691 Peptide #3125 encoded by probe for

1: KKKKK (R,K){20,20} (K){20} KKKKK
2: K KKKKK (K){20} KKKKK
3: KK KKKKK (K){20} KKKKK
4: KKK KKKKK (K){20} KKKKK
5: KKKK KKKKK (K){20} KKKKK
6: KKKKK KKKKK (K){20} KKKKK
7: KKKKK KKKKK (K){20} KKKKK
8: KKKKK KKKKK (K){20} KKKKK
9: KKKKK KKKKK (K){20} KKKKK

1

10: KKKKK (K){20} KKKKK
11: KKKKK (K){20} KKKKK
12: KKKKK (K){20} KKKKK
13: KKKKK (K){20} KKKKK
14: KKKKK (K){20} KKKKK
15: KKKKK (K){20} KKKKK
16: KKKKK (K){20} KKKKK
17: KKKKK (K){20} KKKKK
18: KKKKK (K){20} KKKKK
19: KKKKK (K){20} KKKKK
20: KKKKK (R,K){20,20} (K){20} KKKKK
21: KKKKK (K){20} KKKKK
22: KKKKK (K){20} KKKKK
23: KKKKK (K){20} KKKKK
24: KKKKK (K){20} KKKKK
25: KKKKK (R,K){20} KKKKK
26: KKKKK (R,K){20} KKKKK
27: KKKKK (R,K){20} KKKKK
28: KKKKK (R,K){20} KKKKK
29: KKKKK (R,K){20} KKKKK
30: KKKKK (R,K){20} KKKKK
31: KKKKK (R,K){20} KKKKK
32: KKKKK (R,K){20} KKKKK
33: KKKKK (R,K){20} KKKKK
34: KKKKK (R,K){20} KKKKK
35: KKKKK (R,K){20} KKKKK
36: KKKKK (R,K){20} KKKKK
37: KKKKK (R,K){20} KKKKK
38: KKKKK (R,K){20} KKKKK
39: KKKKK (R,K){20} KKKKK
40: KKKKK (R,K){20} KKKKK

AAM17388 ck: 2686 len: 71 | Aam17388 Peptide #3822 encoded by probe for

1

41: KKKKK (R,K){20} KKKKKK KKKKK
 42: KKKKK (R,K){20} KKKKKK KKKKK
 43: KKKKK (R,K){20} KKKKKK KKKKK
 44: KKKKK (R,K){20} KKKKKK KKKKK
 45: KKKKK (R,K){20} KKKKKK KKKKK
 46: KKKKK (R,K){20} KKKKKK KKKKK
 47: KKKKK (K){20} KKKKKK KKKKK
 48: KKKKK (K){20} KKKKKK KKKKK
 49: KKKKK (K){20} KKKKKK KKKKK
 AAM18532 ck: 1560 len: 88 ! Aam18532 Peptide #4966 encoded by probe for
 43: RRRRG (R,K){20,20} RRRRRR RRRRR
 44: RRRRG (R){20} RRRRRR RRRRR
 45: RRRRG (R){20} RRRRRR RRRRR
 46: RRRRG (R){20} RRRRRR RRRRR
 47: RRRRG (R){20} RRRRRR RRRRR
 48: RRRRG (R){20} RRRRRR RRRRR
 49: RRRRG (R){20} RRRRRR RRRRR
 50: RRRRG (R){20} RRRRRR RRRRR
 51: RRRRG (R){20} RRRRRR RRRRR
 52: RRRRG (R){20} RRRRRR RRRRR
 53: RRRRG (R){20} RRRRRR RRRRR
 54: RRRRG (R){20} RRRRRR RRRRR
 55: RRRRG (R){20} RRRRRR RRRRR
 56: RRRRG (R){20} RRRRRR RRRRR

57: RRRRR (R){20} RRRRRR RRRRR
 58: RRRRR (R){20} RRRRRR RRRRR
 59: RRRRR (R){20} RRRRRR RRRRR
 60: RRRRR (R){20} RRRRRR RRRRR
 61: RRRRR (R){20} RRRRRR RRRRR
 62: RRRRR (R){20} RRRRRR RRRRR
 AAM20089 ck: 8343 len: 66 ! Aam20089 Peptide #6523 encoded by probe for
 6: ETERE (R,K){20,20} (R,K){20} KKKKK
 7: TEREX (R,K){20} KKKKKK KKKKK
 8: EREKK (R,K){20} KKKKKK KKKKK
 9: REKKK (R,K){20} KKKKKK KKKKK
 AAM21941 ck: 4895 len: 51 ! Aam21941 Peptide #8375 encoded by probe for
 18: LFKPM (R,K){20,20} (R,K){20} RRRRRR RRRRR
 19: PKPMR (R,K){20} RRRRRR RRRRR
 20: KPMRK (R,K){20} RRRRRR RRRRR
 AAM26526 ck: 5383 len: 86 ! Aam26526 Peptide #563 encoded by probe for
 15: RRRRG (R,K){20,20} (R,K){20} RRRRRR RRRRR
 16: RRRRG (R,K){20} RRRRRR RRRRR
 17: RRRRG (R,K){20} RRRRRR RRRRR
 18: RRRRG (R,K){20} RRRRRR RRRRR
 19: RRRRG (R,K){20} RRRRRR RRRRR
 20: RRRRG (R,K){20} RRRRRR RRRRR
 21: RRRRK (R,K){20} KKKKKK RRRRR

43: GRKR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 44: RKRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 45: KRRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 46: ERRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 47: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 48: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 49: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 50: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 51: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 52: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 53: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR
 54: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR RRRR

AAM28616 ck: 3607 len: 88 | Aam28616 Peptide #2653 encoded by probe for

39: ERKR (R,K){20,20} (K){20} RRRRRRRRRRRRRRRRRR RRRR
 40: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 41: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 42: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 43: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 44: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 45: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 46: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 47: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 48: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 49: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR

50: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 51: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 52: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 53: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 54: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 55: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 56: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 57: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 58: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 59: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 60: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 61: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 62: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 63: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 64: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 65: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 66: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 67: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 68: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 69: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR

AAM29178 ck: 3937 len: 85 | Aam29178 Peptide #3215 encoded by probe for

1: RRRR (R,K){20,20} (K){20} RRRRRRRRRRRRRRRRRR RRRR
 2: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR
 3: RRRR (K){20} RRRRRRRRRRRRRRRRRR RRRR

```

1:  (R, K) {20}
   RRRRRRKKKKRRRRRRRR RRKT

2:  (R, K) {20}
   R RRRRRRKKKKRRRRRRRR RRKT

```


3: RR RRRRRKKKKRRRRRRRRK RT
 4: RRR RRRKKKKRRRRRRRRR R T

AAM37794 ck: 3301 len: 52 I Aam37794 Peptide #11831 encoded by probe fd

12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AAM38273 ck: 4895 len: 51 I Aam38273 Peptide #12310 encoded by probe fd

18: LFKPM RRRRRKKKKRRRRRRRRR KKLTT
 19: FKPMR RRRRRKKKKRRRRRRRRR KLT TT
 20: KPMRK RRRRRKKKKRRRRRRRRR LTTT

AAU04283 ck: 8137 len: 45 I Aau04283 Trimeric fusogenic peptide #2 use

4: YKA KKKKKKKKKKKKKKKKKKK KKKKK
 5: YKAK KKKKKKKKKKKKKKKKKKK KKKKK
 6: YKAKK KKKKKKKKKKKKKKKKKKK KKKKK
 7: KAKKK KKKKKKKKKKKKKKKKKKK KKKKK
 8: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

23: KKKKK KKKKKKKKKKKKKKKKKKK KWK
 (K) {20}
 24: KKKKK KKKKKKKKKKKKKKKKKKK KWK
 (K) {20}

AAU04285 ck: 4361 len: 59 ! Aau04285 Nuclear ligand #2 used in nucleic
 (R,K) {20,20}
 18: APYKA KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 19: PYKAK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 20: YKAKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 21: KAKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 22: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 32: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 37: KKKKK KKKKKKKKKKKKKKKKKKK KWK
 (K) {20}
 38: KKKKK KKKKKKKKKKKKKKKKKKK KWK
 (K) {20}

1

1

AAU04287 ck: 4925 len: 100 ! Aau04287 Poly-L-Lysine used in nucleic acid
 (R,K) {20,20}
 1: KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 2: K KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

[illegible]

49:	KKKK	{K} {20}	KKKK
50:	KKKK	{K} {20}	KKKK
51:	KKKK	{K} {20}	KKKK
52:	KKKK	{K} {20}	KKKK
53:	KKKK	{K} {20}	KKKK
54:	KKKK	{K} {20}	KKKK
55:	KKKK	{K} {20}	KKKK
56:	KKKK	{K} {20}	KKKK
57:	KKKK	{K} {20}	KKKK
58:	KKKK	{K} {20}	KKKK
59:	KKKK	{K} {20}	KKKK
60:	KKKK	{K} {20}	KKKK
61:	KKKK	{K} {20}	KKKK
62:	KKKK	{K} {20}	KKKK
63:	KKKK	{K} {20}	KKKK
64:	KKKK	{K} {20}	KKKK
65:	KKKK	{K} {20}	KKKK
66:	KKKK	{K} {20}	KKKK
67:	KKKK	{K} {20}	KKKK
68:	KKKK	{K} {20}	KKKK
69:	KKKK	{K} {20}	KKKK
70:	KKKK	{K} {20}	KKKK
71:	KKKK	{K} {20}	KKKK
72:	KKKK	{K} {20}	KKKK

63: RRRK (R,K){20} KRRKKKKKKKKKKKKKKKK
64: RRRK (R,K){20} RRRKKKKKKKKKKKKKKKK
65: RRRK (R,K){20} RRRKKKKKKKKKKKKKKKK
66: RRRK (R,K){20} RRRKKKKKKKKKKKKKKKK
67: RRRK (R,K){20} RRRKKKKKKKKKKKKKKKK

AAM02768 ck: 9082 len: 167 ! Aam02768 Peptide #1450 encoded by probe for

1

33: EGRG (R,K){20,20} RRRRRRRRRRRRRRRRRR
34: EGRG (R,K){20} RRRRRRRRRRRRRRRRRR
35: GRGR (R,K){20} RRRRRRRRRRRRRRRRRR

AAM03278 ck: 2276 len: 89 ! Aam03278 Peptide #1960 encoded by probe for

1

23: EEEE (R,K){20,20} KRRKKKKKKKKKKKKKKKK
46: KEEE (R,K){20} KRRKKKKKKKKKKKKKKKK
47: KEEK (R,K){20} KRRKKKKKKKKKKKKKKKK
48: EEEK (R,K){20} KRRKKKKKKKKKKKKKKKK
49: EEEK (R,K){20} KRRKKKKKKKKKKKKKKKK

AAM03564 ck: 1939 len: 130 ! Aam03564 Peptide #2246 encoded by probe for

1

42: EGRK (R,K){20,20} RRRRRRRRRRRRRRRRRR
43: GRKR (R,K){20} RRRRRRRRRRRRRRRRRR
44: RKRR (R,K){20} RRRRRRRRRRRRRRRRRR
45: KERR (R,K){20} RRRRRRRRRRRRRRRRRR
46: ERRR (R,K){20} RRRRRRRRRRRRRRRRRR
47: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR
48: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR

49: RRRR RRRRRRRRRRRRRRRRRR
50: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR
51: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR
52: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR
53: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR
54: RRRR (R,K){20} RRRRRRRRRRRRRRRRRR

AAM03848 ck: 3607 len: 88 ! Aam03848 Peptide #2530 encoded by probe for

1

39: ERRE (R,K){20,20} KRRKKKKKKKKKKKKKKKK
40: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
41: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
42: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
43: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
44: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
45: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
46: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
47: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
48: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
49: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
50: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
51: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
52: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
53: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
54: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK
55: RREK (R,K){20} KRRKKKKKKKKKKKKKKKK

56: KKKK KKKK {K} {20} KKKK
57: KKKK KKKK {K} {20} KKKK
58: KKKK KKKK {K} {20} KKKK
59: KKKK KKKK {K} {20} KKKK
60: KKKK KKKK {K} {20} KKKK
61: KKKK KKKK {K} {20} KKKK
62: KKKK KKKK {K} {20} KKKK
63: KKKK KKKK {K} {20} KKKK
64: KKKK KKKK {K} {20} KKKK
65: KKKK KKKK {K} {20} KKKK
66: KKKK KKKK {K} {20} KKKK
67: KKKK KKKK {K} {20} KK
68: KKKK KKKK {K} {20} K
69: KKKK KKKK {K} {20} KKKK

AAM04408 ck: 3937 len: 85 ! Aam04408 Peptide #3090 encoded by probe for

1:	$(R, K) \{20, 20\}$ (K) {20}	XXXXXXXXXXXXXXXXXXXX	XXXXX
2:	K	XXXXXXXXXXXXXXXXXXXX	XXXXXXXX
3:	KK	(K) {20}	XXXXXXXXXXXX
4:	KKK	(K) {20}	XXXXXXXXXXXX
5:	KKKK	(K) {20}	XXXXXXXXXXXX
6:	KKKKK	(K) {20}	XXXXXXXXXXXX
7:	KKKKKK	(K) {20}	XXXXXXXXXXXX
8:	KKKKKKK	(K) {20}	XXXXXXXXXXXX
9:	KKKKKKKK	(K) {20}	XXXXXXXXXXXX

[illegible]

AAG73687 ck: 3063 len: 29 ! Aag73687 Human colon cancer antigen protein
(R,K){20,20}

8: MMWTF KKKKKKKKKKKKKKKKKKKKKKK KX
9: MMWTFK KKKKKKKKKKKKKKKKKKKKKKK X
AAG73729 ck: 783 len: 83 ! Aag73729 Human colon cancer antigen protein
(K) {20}
(R, K) {20, 20}
49: LGPCE KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
50: GPCEK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
51: PCEKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
52: CEKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
53: EKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
AAG73810 ck: 3374 len: 88 ! Aag73810 Human colon cancer antigen protein
(R, K) {20, 20}
(K) {20}
44: FGQTX KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
45: GQTXK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
46: QTXKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
47: TXKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
48: XKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
50: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
52: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
53: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

(K) {20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
60: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
61: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
AAG73895 ck: 1887 len: 43 ! Aag73895 Human colon cancer antigen protein
(R, K) {20, 20}
(R, K) {20}
18: VRPRV RKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
19: RPRVR KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
20: PRVRK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
21: RVRKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
22: VRKKK KKKKKKKKKKKKKKKKKKKKKKK GG
(K) {20}
AAG74218 ck: 8659 len: 104 ! Aag74218 Human colon cancer antigen protein
(R, K) {20, 20}
(K) {20}
75: PLGGQ KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
76: LGGQK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
77: GGQKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
78: GQKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
79: QKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}
80: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K) {20}

[illegible]

AAG74650 ck: 1596 len: 69 ! Aag74650 Human colon cancer antigen protein

[illegible]

AAG74793 ck: 8497 len: 152 ! Aag74793 Human colon cancer antigen protein
(R,K) { 20, 20 }
(K) { 20 }
22: SSHTQ KKKKKKKKKKKKKKKKKKKKKK KKKKK

[illegible]

1 AAG75215 ck: 3913 le
 $(R, K) \{20, 20\}$
 $(K) \{20\}$

135: RSSAP KKKKKKKKKKKKKKKKKKKKK K
 (K) {20}
 136: SSAPK KKKKKKKKKKKKKKKKKKKKK
 AAG75886 ck: 4235 len: 71 ! Aag75886 Human colon cancer antigen protein
 (R,K) {20,20}
 47: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGXX
 (K) {20}
 48: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGX
 (K) {20}
 AAE01796 ck: 4416 len: 72 ! Aae01796 Human gene 27 encoded secreted pro
 (R,K) {20,20}
 47: LPTFL KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 48: PTFLK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 49: TFLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 50: FLKKK KKKKKKKKKKKKKKKKKKKKK KKK
 (K) {20}
 51: LKKKK KKKKKKKKKKKKKKKKKKKKK KK
 (K) {20}
 52: KKKKK KKKKKKKKKKKKKKKKKKKKK K
 (K) {20}
 53: KKKKK KKKKKKKKKKKKKKKKKKKKK
 AAE01848 ck: 5584 len: 73 ! Aae01848 Human gene 27 encoded secreted pro
 (R,K) {20,20}
 47: LPTFL KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 48: PTFLK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 49: TFLKK KKKKKKKKKKKKKKKKKKKKK KKKKI
 (K) {20}
 50: FLKKK KKKKKKKKKKKKKKKKKKKKK KKKI
 (K) {20}
 51: LKKKK KKKKKKKKKKKKKKKKKKKKK KKI
 (K) {20}
 52: KKKKK KKKKKKKKKKKKKKKKKKKKK KI
 (K) {20}
 53: KKKKK KKKKKKKKKKKKKKKKKKKKK I

1
 AAB45846 ck: 8137 len: 45 ! Aab45846 Nucleic acid transporter system pr
 (R,K) {20,20}
 4: YKA KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 5: YKAK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 6: YKAKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 7: KAKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 8: AKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}
 24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K) {20}

1
 AAB90574 ck: 1431 len: 530 ! Aab90574 Human secreted protein, SEQ ID NO:
 (R,K) {20,20}
 511: LHAPP KKKKKKKKKKKKKKKKKKKKK

1
 AAB45848 ck: 4361 len: 59 ! Aab45848 Nucleic acid transporter system pr
 (R,K) {20,20}
 18: APYKA KKKKKKKKKKKKKKKKKKKKK KKKKK

19: PYAKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
20: YKAKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
21: KAKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
22: AKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
23: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
24: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
25: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
26: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
27: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
28: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
29: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
30: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
31: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
32: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
33: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
34: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
35: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
36: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
37: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
38: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK

AAB45850 ck: 4925 len: 100 ! Aab45850 Nucleic acid transporter system pe

1: (R,K) {20,20} KKKKKK KKKKKK KKKKKK KKKKKK
2: K KKKKKK KKKKKK KKKKKK KKKKKK
3: KK KKKKKK KKKKKK KKKKKK KKKKKK

4: KKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
5: KKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
6: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
7: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
8: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
9: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
10: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
11: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
12: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
13: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
14: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
15: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
16: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
17: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
18: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
19: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
20: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
21: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
22: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
23: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
24: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
25: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
26: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK
27: KKKKK (K) {20} KKKKKK KKKKKK KKKKKK KKKKKK

77: KKKK (K){20} KKKK
78: KKKK (K){20} KKKK
79: KKKK (K){20} KKKK
80: KKKK (K){20} KKKK
81: KKKK (K){20} KKKK

AAB50247 ck: 8085 len: 154 ! Aab50247 Human breast cancer associated B72

(R,K){20,20}
(K){20}

114: TQLRQ KKKK KKKK KKKK KKKK KKKK
115: QLROK KKKK KKKK KKKK KKKK KKKK
116: LRQKK KKKK KKKK KKKK KKKK KKKK
117: RQKKK KKKK KKKK KKKK KKKK KKKK
118: QKKKK KKKK KKKK KKKK KKKK KKKK
119: KKKKK KKKK KKKK KKKK KKKK KKKK
120: KKKKK KKKK KKKK KKKK KKKK KKKK
121: KKKKK KKKK KKKK KKKK KKKK KKKK
122: KKKKK KKKK KKKK KKKK KKKK KKKK
123: KKKKK KKKK KKKK KKKK KKKK KKKK
124: KKKKK KKKK KKKK KKKK KKKK KKKK
125: KKKKK KKKK KKKK KKKK KKKK KKKK
126: KKKKK KKKK KKKK KKKK KKKK KKKK
127: KKKKK KKKK KKKK KKKK KKKK KKKK
128: KKKKK KKKK KKKK KKKK KKKK KKKK

ABJ37738 ck: 8085 len: 154 ! Abj37738 Human tumour-related protein - SEQ

(R,K){20,20}
(K){20}

114: TQLRQ KKKK KKKK KKKK KKKK KKKK
115: QLROK KKKK KKKK KKKK KKKK KKKK

116: LRQKK KKKK (K){20} KKKK
117: RQKKK KKKK (K){20} KKKK
118: QKKKK KKKK (K){20} KKKK
119: KKKKK KKKK (K){20} KKKK
120: KKKKK KKKK (K){20} KKKK
121: KKKKK KKKK (K){20} KKKK
122: KKKKK KKKK (K){20} KKKK
123: KKKKK KKKK (K){20} KKKK
124: KKKKK KKKK (K){20} KKKK
125: KKKKK KKKK (K){20} KKKK
126: KKKKK KKKK (K){20} KKKK
127: KKKKK KKKK (K){20} KKKK
128: KKKKK KKKK (K){20} KKKK

ABR00951 ck: 6732 len: 139 ! ABR00951 Human gene 5-encoded secreted prot

1
(R,K){20,20}
(K){20}

91: HPRPG KKKK KKKK KKKK KKKK KKKK
92: FRPGK KKKK (K){20} KKKK
93: RPKGK KKKK (K){20} KKKK
94: PGKKK KKKK (K){20} KKKK
95: GKKKK KKKK (K){20} KKKK
96: KKKKK KKKK (K){20} KKKK
97: KKKKK KKKK (K){20} KKKK
98: KKKKK KKKK (K){20} KKKK
99: KKKKK KKKK (K){20} KKKK
100: KKKKK KKKK (K){20} KKKK

20: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
21: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
22: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
23: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
24: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
25: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
26: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
27: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
28: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
29: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
30: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
31: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
32: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
33: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
34: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
35: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
36: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
37: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}

ABG73872 ck: 9643 len: 83 ! Abg73872 Single-chain antigen-binding (sFv)

1

1: (R,K){20,20}
(R,K){20}
R RRRRRRRRRRRRRRRR RRRR
2: R RRRRRRRRRRRRRRRR RRRR
(R,K){20}
3: R RRRRRRRRRRRRRRRR RRRR
(R,K){20}
4: R RRRRRRRRRRRRRRRR RRRR
(R,K){20}
5: R RRRRRRRRRRRRRRRR RRRR
(R,K){20}

6: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
7: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
8: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
9: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
10: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
11: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
12: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
13: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
14: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
15: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
16: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
17: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
18: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
19: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
20: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
21: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
22: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
23: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
24: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
25: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
26: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
27: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
28: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
29: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
30: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}


```

354: KKKKK (K) {20}
ABJ19654 ck: 6732 len: 139 ! Abj
(R,K) {20,20}
91: HFRPG KKKKK (K) {20} KKKKK
92: FRPGK KKKKK (K) {20} KKKKK
93: RPKGK KKKKK (K) {20} KKKKK
94: PGKKK KKKKK (K) {20} KKKKK
95: GKKKK KKKKK (K) {20} KKKKK
96: KKKKK KKKKK (K) {20} KKKKK
97: KKKKK KKKKK (K) {20} KKKKK
98: KKKKK KKKKK (K) {20} KKKKK
99: KKKKK KKKKK (K) {20} KKKKK
100: KKKKK KKKKK (K) {20} KKKKK
101: KKKKK KKKKK (K) {20} KKKKK
102: KKKKK KKKKK (K) {20} KKKKK
103: KKKKK KKKKK (K) {20} KKKKK
104: KKKKK KKKKK (K) {20} KKKKK
105: KKKKK KKKKK (K) {20} KKKKK
106: KKKKK KKKKK (K) {20} KKKKK
107: KKKKK KKKKK (K) {20} KKKKK
108: KKKKK KKKKK (K) {20} KKKKK
109: KKKKK KKKKK (K) {20} KKKKK
110: KKKKK KKKKK (K) {20} KKKKK
111: KKKKK KKKKK (K) {20} KKKKK
112: KKKKK KKKKK (K) {20} KKKKK

```

113: KKKKK (K) {20} KKKKK
 114: KKKKK (K) {20} KKKKK
 115: KKKKK (K) {20} KKKKK
 116: KKKKK (K) {20} KKKKK
 117: KKKKK (K) {20} KKKKK
 118: KKKKK (K) {20} KKKKK
 119: KKKKK (K) {20} KKKKK
 120: KKKKK (K) {20} KKKKK

ABP99475 ck: 6732 len: 139 ! Abp99475 Human secreted protein SEQ ID NO 4

91: HFRPG (R,K) {20,20} KKKKK
 92: FRPGK (K) {20} KKKKK
 93: RPKKK (K) {20} KKKKK
 94: PPKKK (K) {20} KKKKK
 95: GKKKK (K) {20} KKKKK
 96: KKKKK (K) {20} KKKKK
 97: KKKKK (K) {20} KKKKK
 98: KKKKK (K) {20} KKKKK
 99: KKKKK (K) {20} KKKKK
 100: KKKKK (K) {20} KKKKK
 101: KKKKK (K) {20} KKKKK
 102: KKKKK (K) {20} KKKKK
 103: KKKKK (K) {20} KKKKK
 104: KKKKK (K) {20} KKKKK
 105: KKKKK (K) {20} KKKKK

106: KKKKK (K) {20} KKKKK
 107: KKKKK (K) {20} KKKKK
 108: KKKKK (K) {20} KKKKK
 109: KKKKK (K) {20} KKKKK
 110: KKKKK (K) {20} KKKKK
 111: KKKKK (K) {20} KKKKK
 112: KKKKK (K) {20} KKKKK
 113: KKKKK (K) {20} KKKKK
 114: KKKKK (K) {20} KKKKK
 115: KKKKK (K) {20} KKKKK
 116: KKKKK (K) {20} KKKKK
 117: KKKKK (K) {20} KKKKK
 118: KKKKK (K) {20} KKKKK
 119: KKKKK (K) {20} K
 120: KKKKK (K) {20} KKKKK

ABP99639 ck: 9531 len: 66 ! Abp99639 Human secreted protein SEQ ID NO

41: MWTVX (R,K) {20,20} KKKKK
 42: WTVXK (K) {20} KKKKK
 43: TVXKK (K) {20} KKKKK
 44: VXXKK (K) {20} KKKKK
 45: XXXKK (K) {20} KKKKK
 46: KKKKK (K) {20} K
 47: KKKKK (K) {20} KKKKK

Databases searched:

Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

Total finds: 8,655
Total length: 158,726,570
Total sequences: 1,107,863
CPU time: 07:06.66
